CAPÍTULO 4

Problemas estadísticos de dos muestras y más

En los §§ 1 y 2 se examinan los problemas de homogeneidad de dos muestras.

En el § 3 se estudian los problemas de regresión.

En el § 4 se exponen los resultados del análisis de varianza.

En el § 5 se examinan los problemas de reconocimiento de las imágenes.

§ 1. Verificación de las hipótesis de homogeneidad (completa o parcial) en el caso paramétrico

1. Clase de problemas a examinar. En los capítulos anteriores, el objeto de todos los estudios ha sido la muestra X de volumen n de una distribución $\mathbf{P}$ total o parcialmente desconocida. Ahora pasamos al estudio de los problemas estadísticos donde figura no una, *sino dos muestras y más.*

Una de las clases principales de problemas que se examinan en este caso son los problemas de verificación de la homogeneidad (completa o parcial) de dos muestras.

Aquí entran los tres siguientes tipos principales de problemas:

I. *Verificación de la homogeneidad "ordinaria".* Aquí el problema consiste en verificar la hipótesis de que dos muestras X e Y se han extraído de una misma distribución desconocida. Tales problemas surgen, por ejemplo, al comparar dos métodos de elaboración en cualquier proceso tecnológico o en la agricultura. Como base de comparación suelen servir las características numéricas del producto final (de la muestra), que son de naturaleza aleatoria. Problemas de este mismo género surgirán si por el estado de salud de los enfermos verificamos el efecto de una nueva medicina, comparando el grupo experimental de pacientes con el grupo de control.

Entre los problemas de homogeneidad figura el ejemplo dado en la introducción.

En este párrafo examinaremos el *caso paramétrico.* Supongamos que se da una familia de distribuciones $\{\mathbf{P}_\theta\}_{\theta\in\Theta}$ y que hay dos muestras independientes $X = (x_1, ..., x_{n_1})$ e $Y = (y_1, ..., y_{n_2})$ de volúmenes n_1 y n_2, respectivamente, con la particularidad de que se sabe de antemano que estas muestras pertenecen a la familia $\{\mathbf{P}_\theta\}$:

$$X \in \mathbf{P}_{\theta_1}, \qquad Y \in \mathbf{P}_{\theta_2} \tag{1}$$

para ciertos θ_1 y θ_2. El problema ordinario de homogeneidad aquí consiste en verificar la hipótesis $H_1 = \{\theta_1 = \theta_2\}$ frente a la alternativa adicional $H_2 = \{\theta_1 \neq \theta_2\}$. Es evidente que aquí ambas hipótesis H_1 y H_2 son compuestas.

II. *Verificación de la homogeneidad al existir un parámetro obstaculizador.* Aquí se supone que la dimensión k del parámetro θ es mayor que 1. Escribamos el vector θ en forma de la colección $\theta = (u, v)$ de dos subvectores u y v y designemos por u_j las componentes de los vectores θ_j en (1), $j = 1, 2$.

Supongamos que sabemos de antemano que en ambas muestras, "el subparámetro", a pesar de ser desconocido, es común: $v_1 = v_2 = v$. Se verifica la hipótesis $H_1 = \{u_1 = u_2\}$ frente a $H_2 = \{u_1 \neq u_2\}$.

Este es precisamente un problema de homogeneidad cuando se dispone del parámetro obstaculizador v. El mismo se distingue de los problemas ordinarios de homogeneidad por el hecho de que la alternativa para la hipótesis $H_1 = \{\theta_1 = \theta_2\}$ tiene la forma $H_2 = \{u_1 \neq u_2,\ v_1 = v_2\}$.

Se puede citar el siguiente ejemplo de surgimiento de tal tipo de problemas. Supongamos que nos interesa el estado de cierto objeto que se caracteriza por el vector a que no puede ser medido directamente. Podemos efectuar tan sólo mediciones en las que sobre a se superpone un ruido aleatorio cuya naturaleza, al efectuar diversas observaciones, permanece invariable. Debemos verificar la hipótesis de invariabilidad de a en dos series de observaciones X e Y.

Si, digamos, las mediciones tienen la forma $x_i = a_1 + \xi_i$, donde $\xi_i \in \Phi_{\lambda,\sigma^2}$ determinan el papel que desempeña el ruido, y las observaciones y_i tienen ese mismo carácter al sustituir a_1 por a_2, entonces podemos escribir $X \in \Phi_{a_1+\lambda,\sigma^2}$, $Y \in \Phi_{a_2+\lambda,\sigma^2}$. Hemos llegado al problema de verificación de la igualdad de las medias $\{\alpha_1 = \alpha_2\}$ de dos poblaciones normales Φ_{α_1,σ^2} y Φ_{α_2,σ^2} para el valor desconocido común σ^2.

III. *Verificación de la homogeneidad parcial.* Aquí solamente se verifica la hipótesis H_1 acerca de la coincidencia "parcial" de θ_1 y θ_2. Es decir, se comprueba la hipótesis $H_1 = \{u_1 = u_2\}$ (con designaciones del apartado anterior) frente a $H_2 = \{u_1 \neq u_2\}$. Los valores de v_1 y v_2 pueden ser propios para cada una de las muestras X e Y.

Supongamos, por ejemplo, que en un laboratorio se estima el resultado de la influencia que ejerce un nuevo método de cultivo sobre el rendimiento de cualquier cereal. Las observaciones representan el peso total de los granos en distintas espigas. Supongamos que $x_i \in \Phi_{\alpha_1,\sigma_1^2}$, $i = 1, ..., n_1$ para una partida experimental de espigas, e $y_i \in \Phi_{\alpha_2,\sigma_2^2}$ para la partida de control. Es natural admitir que la "dispersión" σ^2 puede variar a consecuencia del cambio de cultivo. Pero para nosotros es importante saber si cambia o no el índice principal α que determina el rendimiento del cereal. Llegamos al problema de verificación de la hipótesis $H_1 = \{\alpha_1 = \alpha_2\}$ frente a $H_2 = \{\alpha_1 \neq \alpha_2\}$ para poblaciones normales cuyas varianzas pueden ser diferentes. En la literatura, este problema es conocido con el nombre de *problema de Behrens — Fisher**).

En este párrafo reduciremos los problemas de todos los tres tipos, para las familias paramétricas arbitrarias, al problema examinado en el § 3.15, de pertenencia de una muestra a una subfamilia paramétrica, y hallaremos una serie de criterios asintóticamente minimax, suponiendo la semejanza de las hipótesis sometidas a verificación. Serán los criterios de relación de verosimilitud que, para poblaciones normales, coincidirán con los criterios construidos al buscar una u otra optimización exacta (si tales existen; compárese con [57]).

El criterio estadístico π para verificar H_1 frente a H_2, en nuestro caso será la función $\pi = \pi(X, Y)$ de dos muestras X e Y que, al igual que en la exposición anterior, designará la probabilidad de aceptación de H_2 para una muestra unida dada (X, Y) (véase el capítulo 3). Las definiciones del nivel asintótico y de la optimización asintótica del criterio π aquí son las mismas que en el § 3.14.

Definición 1. Diremos que el criterio π *tiene un nivel asintótico* $1 - \varepsilon$ (pertenece a la clase $\bar{K}_\varepsilon$), si

$$\lim_{n\to\infty}\sup \sup_{(\theta_1,\theta_2)\in\bar{\Theta}_1} \mathbf{M}_{\theta_1\theta_2}\pi(X, Y) \leqslant \varepsilon_1,$$

donde $\mathbf{M}_{\theta_1\theta_2}$ significa la esperanza matemática respecto a la distribución $\mathbf{P}_{\theta_1} \times \mathbf{P}_{\theta_2}$, y $\bar{\Theta}_1$ es el conjunto de valores (θ_1, θ_2) con los que se cumple

*) Se han escrito muchos libros dedicados a la búsqueda de sus soluciones óptimas. Al estudio del problema de Behrens — Fisher, que resultó muy difícil, contribuyeron considerablemente las investigaciones de Yu. V. Línnik y sus alumnos. Dichas investigaciones requieren la introducción de nuevos conceptos y el uso de un aparato matemático muy complejo. Esto hace imposible la enunciación y demostración (en el marco de este manual) de los resultados obtenidos. La situación acerca de los problemas de homogeneidad ordinaria y de homogeneidad para poblaciones normales al existir un parámetro obstaculizador, es algo mejor (en una serie de problemas se logra hallar los criterios invariantes no desplazados y uniformemente más potentes). No obstante, las construcciones indispensables para ello también resultan muy complicadas; este tema se examina más detalladamente en [57].

la hipótesis H_1 (por ejemplo, el conjunto de todos los puntos (θ_1, θ_2) situados en la "bisectriz" $\theta_1 = \theta_2$ en el problema de homogeneidad ordinaria).

Definición 2. El criterio $\pi_1 \in \tilde{K}_\varepsilon$ se llama asintóticamente minimax en $\tilde{K}_\varepsilon$ para verificar H_1 frente a H_2, si para cualquier criterio $\pi \in \tilde{K}_\varepsilon$ se cumple

$$\varliminf_{n\to\infty}\left(\inf_{(\theta_1,\theta_2)\in\bar{\Theta}_2} \mathbf{M}_{\theta_1\theta_2}\pi_1(X,\ Y) - \inf_{(\theta_1,\theta_2)\in\bar{\Theta}_2} \mathbf{M}_{\theta_1\theta_2}\pi(X,\ Y)\right) \geqslant \theta_1,$$

donde $\bar{\Theta}_2$ es el conjunto de valores (θ_1, θ_2) correspondientes a las alternativas de H_2.

2. Criterio asintóticamente minimax para verificar las hipótesis semejantes de homogeneidad ordinaria. Introduzcamos un nuevo parámetro $\bar{\theta} = (\theta_1, \theta_2)$ que caracterice la muestra "unida" (X, Y). La función de verosimilitud de la muestra es igual a $f_{\bar{\theta}}(X,\ Y) = f_{\theta_1}(X)f_{\theta_2}(Y)$.

Supongamos primeramente, para abreviar, que los volúmenes de las muestras coinciden: $n_1 = n_2 = n$. Entonces, la muestra $(X,\ Y)$ puede representarse como muestra de volumen n formada por las observaciones $(x_1, y_1), \ldots, (x_n, y_n)$ de la distribución $\mathbf{P}_{\bar{\theta}} = \mathbf{P}_{\theta_1} \times \mathbf{P}_{\theta_2}$ que tiene la densidad $f_{\theta_1}(x)f_{\theta_2}(y)$. Llegamos al problema examinado en el § 3.15, de verificación, a base de la muestra $(X,\ Y)$, de la hipótesis H_1 de que el parámetro $\bar{\theta}$ se sitúa en la "curva" $\theta_1 = \theta_2$. Teniendo en cuenta las designaciones de § 3.15, en nuestro caso, la hipótesis H_1 tiene la forma $\bar{\theta} = g(\alpha)$, donde $\alpha \equiv \theta_1$, $g(\alpha) = (\alpha,\ \alpha)$. Es evidente que la matriz $G = \left\| \dfrac{\partial g_i}{\partial \alpha_j} \right\|$, $i = 1, \ldots$ $\ldots, 2k$, $j = 1, \ldots, k$, tiene la forma $\begin{pmatrix} E \\ E \end{pmatrix}$, donde E es la matriz unidad de k-ésimo orden, así que el rango de G es igual a k.

Consideraremos localizado el parámetro $\bar{\theta}$, o sea, consideraremos que los valores de θ_1 y θ_2 son semejantes y, por consiguiente, que los posibles valores de $\bar{\theta}$ se sitúan en el entorno del punto $\bar{\theta}_0 = (\theta_0, \theta_0)$ para cierto θ_0 registrado. Si seguimos el § 3.15, nos será más cómodo introducir un nuevo parámetro $\tau = (\tau',\ \tau'') = (\gamma'/\sqrt{n},\ \gamma''/\sqrt{n}) = \gamma/\sqrt{n}$, donde $\tau' = \theta_1 - \theta_0$, $\tau'' = \theta_2 - \theta_1$, así que la aplicación $\bar{\theta} = \bar{\theta}(\tau)$ es biunívoca: $\theta_1 = \tau' + \theta_0$, $\theta_2 = \tau'' + \tau' + \theta_0$. En los términos de los parámetros τ y γ, la hipótesis H_1 de homogeneidad tomará la forma $H_1 = \{\tau'' = 0\} = \{\gamma'' = 0\}$. En calidad de alternativa examinaremos la hipótesis "aislada"

$$H_2^b = \{\gamma'' I \gamma''^T \geqslant b^2\},\ b > 0, \tag{2}$$

donde $I = I(\theta_0)$ es la matriz de Fisher para la familia $\{\mathbf{P}_\theta\}$ en el punto θ_0.

Teorema 1. *Supongamos que en el entorno del punto θ_0, la familia $\{\mathbf{P}_\theta\}$ satisface las condiciones* (RR) (véase el § 2.28). *Entonces, el criterio de rela-*

ción de verosimilitud

$$R_1(X,\ Y) = \frac{\sup\limits_{\theta_1,\theta_2} f_{\theta_1}(X) f_{\theta_2}(Y)}{\sup\limits_{\theta} f_\theta(X) f_\theta(Y)} > e^{h_\varepsilon/2} \tag{3}$$

es el criterio asintóticamente minimax de nivel $1 - \varepsilon$ *para verificar* $H_1 = \{\theta_1 = \theta_2\}$ frente a $H_2^b = \{(\theta_1 - \theta_2) I(\theta_1 - \theta_2)^T \geqslant b^2/n\}$ *para cualquier* $b > 0$, *donde* h_ε *es una cuantila de orden* $1 - \varepsilon$ *de la distribución* χ^2 *de* k *grados de libertad (para la hipótesis* H_1, *la estadística* $2 \ln R_1(X,\ Y)$ *tiene tal distribución límite).*

Supongamos que $\hat{\theta}_X^*$, $\hat{\theta}_Y^*$, $\hat{\theta}^*$ *es la e.v.m. del parámetro* $\theta = \theta_1 = \theta_2$, *respectivamente, según las muestras* X, Y, $(X,\ Y)$. *Entonces, el criterio*

$$(\hat{\theta}_X^* - \hat{\theta}^*) I(\hat{\theta}^*)(\hat{\theta}_X^* - \hat{\theta}^*)^T + (\hat{\theta}_Y^* - \hat{\theta}^*) I(\hat{\theta}_Y^* - \hat{\theta}^*)^T > h_\varepsilon/n \tag{4}$$

será asintóticamente equivalente al criterio (3).

Demostración. La afirmación mencionada es el corolario directo del teorema 3.15.4. Sólo debemos aclarar qué representa la matriz de Fisher $\bar{I}(\bar{\theta}_0) = \bar{I}(\theta_0,\ \theta_0)$ para el parámetro "unido" $\bar{\theta} = (\theta_1,\ \theta_2)$, y la matriz $\mathbf{M}_2$ para la familia paramétrica $\{\mathbf{P}_{(\theta_0, \theta_0 + \beta)}\}$ en el punto $\beta = 0$. Tenemos

$$\ln f_{\theta_1}(x) f_{\theta_2}(y) = l(x,\ \theta_1) + l(y,\ \theta_2).$$

Designemos por t_i, $i = 1,\ \ldots,\ 2k$ las coordenadas del vector $\bar{\theta}$. En este caso, si por $\mathbf{M}_{\bar{\theta}}$ se designa la esperanza matemática en la distribución $\mathbf{P}_{\bar{\theta}}$, los elementos $\bar{I}_{ij}(\bar{\theta})$ de matriz $\bar{I}(\theta)$ serán iguales a

$$\bar{I}_{ij}(\bar{\theta}) = \mathbf{M}_{\bar{\theta}} \left(\frac{\partial l(\mathrm{x}_1,\ \theta_1)}{\partial t_i} + \frac{\partial l(\mathrm{y}_1,\ \theta_2)}{\partial t_i} \right) \left(\frac{\partial l(\mathrm{x}_1,\ \theta_1)}{\partial t_j} + \frac{\partial l(\mathrm{y}_1,\ \theta_2)}{\partial t_j} \right).$$

De aquí, en virtud de la independencia de x_1 e y_1, obtenemos

$$\bar{I}(\bar{\theta}) = \begin{pmatrix} I(\theta_1) & 0 \\ 0 & I(\theta_2) \end{pmatrix}.$$

Por eso, el criterio (4) no es otra cosa sino el criterio (3.15.12) en el teorema 3.15.4.

Los cálculos análogos muestran que $\mathbf{M}_2 = I(\theta_0)$, ya que para $\beta = \beta(\beta_1,\ \ldots,\ \beta_k) = 0$,

$$\frac{\partial l(\mathrm{x}_1,\ \theta_0)}{\partial \beta_i} + \frac{\partial l(\mathrm{y}_1,\ \theta_0 + \beta)}{\partial \beta_i} = \frac{\partial l(\mathrm{y}_1,\ \theta_0)}{\partial t_i},\ i = 1,\ \ldots,\ k. \ \triangleleft$$

Observación 1. La afirmación del teorema 1 se ha obtenido suponiendo que $n_1 = n_2$. Sin embargo, esta limitación no tiene absolutamente importancia. Examinemos, por ejemplo, el caso cuando $n_1 \to \infty$, $n_2 \to \infty$, de modo que la relación n_1/n_2 sea igual a un número racional r_1/r_2 (r_1 y r_2 son números enteros arbitrarios registrados, $n_i = nr_i$, $n \to \infty$). Volvamos a introducir el nuevo parámetro $\overline{\theta} = (\theta_1, \theta_2)$ y examinemos la muestra unida (X, Y) como una muestra de volumen n con las observaciones $(x_1, ..., x_{r_1}; y_1, ..., y_{r_2}), (x_{r_1+1}, ..., x_{2r_1}; y_{r_2+1}, ..., y_{2r_2}), ...$ de la distribución

$$\mathbf{P}_{\overline{\theta}} = \underbrace{\mathbf{P}_{\theta_1} \times ... \times \mathbf{P}_{\theta_1}}_{r_1 \text{ veces}} \times \underbrace{\mathbf{P}_{\theta_2} \times ... \times \mathbf{P}_{\theta_2}}_{r_2 \text{ veces}}$$

que depende del parámetro $\overline{\theta}$. La función de verosimilitud otra vez adquirirá la forma

$$f_{\overline{\theta}}(X, Y) = f_{\theta_1}(X) f_{\theta_2}(Y).$$

Si se introduce, como antes, el nuevo parámetro $\tau = (\tau', \tau'') = (\theta_1 - \theta_0, \theta_2 - \theta_1)$ y se pone $\tau = \gamma/\sqrt{n} = (\gamma'/\sqrt{n}, \gamma'/\sqrt{n})$, entonces, el problema sometido a examen consiste en verificar $H_1 = \{\gamma'' = 0\}$ frente a $H_2^b = \{\gamma''\mathbf{M}_2\gamma''^T \geqslant b^2\}$, donde $\mathbf{M}_2$ es la matriz de Fisher para $\mathbf{P}_{(\theta_0, \theta_0+\beta)}$ en el punto $\beta = 0$. Es fácil ver que en nuestro caso $\mathbf{M}_2 = r_2 I(\theta_0)$, así que el conjunto de alternativas conserva su forma (2):

$$H_2^b = \{\gamma'' I \gamma''^T > b^2/r^2\}.$$

La matriz de Fisher $I(\overline{\theta})$ tendrá la forma

$$\begin{pmatrix} r_1 I(\theta_1) & 0 \\ 0 & r_2 I(\theta_2) \end{pmatrix}.$$

Sólo queda utilizar el teorema 3.15.4. Entonces obtendremos la afirmación del teorema 1, en la que el criterio (4) ha de sustituirse por

$$n_1(\hat{\theta}_X^* - \theta^*) I(\theta^*)(\hat{\theta}_X^* - \theta^*)^T + \\ + n_2(\hat{\theta}_Y^* - \theta^*) I(\theta^*)(\hat{\theta}_Y^* - \theta^*) > h_\varepsilon. \tag{5}$$

Con ayuda del teorema 3.15.4 también se puede señalar la potencia asintótica garantizada de los criterios (3) — (5).

La afirmación del teorema también es válida en el caso general cuando $n_1 \to \infty$, $n_2 \to \infty$, $n_1/n_2 \to c$, donde c es un número arbitrario de (0, 1). No obstante, la demostración de este hecho exige consideraciones adicionales.

Observación 2. La afirmación del teorema 1 también será válida si la hipótesis $H_1 = \{\theta_1 = \theta_2\}$ se sustituye por la hipótesis (véanse los capítulos precedentes)

$$H_1^a = \{(\theta_1 - \theta_2) I (\theta_1 - \theta_2)^T \leqslant a^2/n\}, \ 0 < a < b.$$

Observación 3. La forma de criterios asintóticamente minimax en el teorema 1 no depende de θ_0. El valor de θ_0 sólo forma parte de la definición de la hipótesis H_2^b a través de $I = I(\theta_0)$ (véase (2), aunque también sería posible evitar la aparición de θ_0 sustituyendo I en (2) por $I((\theta_1 + \theta_2)/2)$. Esto nos proporcionaría la hipótesis H_2^b ("asintóticamente equivalente" a

$\bar{H}^b{}_2$), para la cual se conserva por completo la afirmación del teorema 3. La aparición del valor θ_0 en (2) se debe a la utilización del método más simple de reducción del referido problema a los resultados del § 3.15.

Ejemplo 1. Supongamos que X e Y son muestras de volúmenes n_1 y n_2 de las distribuciones polinomiales $X \in \mathbf{B}_{\theta_1}$, $Y \in \mathbf{B}_{\theta_2}$, $\theta_i \in R^k$, $\theta_i = (\theta_{i1}, \ldots, \theta_{ik})$, $i = 1, 2$. Los vectores de las frecuencias $\nu = (\nu_1, \ldots, \nu_k)$ y $\mu = (\mu_1, \ldots, \mu_k)$ de aparición de los sucesos $A_1, \ldots, A_k$ (véase el § 2.2) forman las estadísticas suficientes

$$f_{\theta_1}(X) = \prod_{i=1}^{k} \theta_{1i}^{\nu_i}, \quad f_{\theta_2}(Y) = \prod_{i=1}^{k} \theta_{2i}^{\mu_i}.$$

Las e.v.m. tienen la forma $\hat{\theta}^*_X = \nu/n_1$, $\hat{\theta}^*_Y = \mu/n_2$, $\hat{\theta}^* = (\nu + \mu)/(n_1 + n_2)$. La matriz $I(\theta)$ está definida en (3.15.5), así que (veáse (3.16.9))

$$tI(\theta_0)t^T = \sum_{i=1}^{k} \frac{t_i^2}{\theta_{0i}}.$$

Así pues, en virtud del teorema 1 y de la observación 1, el criterio asintóticamente minimax de nivel asintótico $1 - \varepsilon$ para verificar $H_1 = \{\theta_1 = \theta_2\}$ frente a

$$H_2^b = \left\{ \sum_{i=1}^{k} (\theta_{1i} - \theta_{2i})^2/\theta_{0i} \geqslant b^2/n_2 \right\}$$

tiene la forma

$$\ln R_1(X, Y) =$$
$$= \sum_{i=1}^{k} \nu_i \ln \frac{\nu_i}{n_1} + \sum_{i=1}^{k} \mu_i \ln \frac{\mu_i}{n_2} - \sum (\nu_i + \mu_i) \ln \frac{\nu_i + \mu_i}{n_1 + n_2} > \frac{h_\varepsilon}{2},$$

donde h_ε es una cuantila del orden de $1 - \varepsilon$ de la distribución χ^2 con $k - 1$ grados de libertad. De acuerdo con (4) y (5), será asintóticamente equivalente el criterio

$$n_1 \sum_{i=1}^{k} \left(\frac{\nu_i}{n_1} - \frac{\nu_i + \mu_i}{n_1 + n_2} \right)^2 \frac{n_1 + n_2}{\nu_i + \mu_i} +$$
$$+ n_2 \sum_{i=1}^{k} \left(\frac{\mu_i}{n_2} - \frac{\nu_i + \mu_i}{n_1 + n_2} \right)^2 \frac{n_1 + n_2}{\nu_i + \mu_i} =$$
$$= \sum_{i=1}^{k} \left(\frac{\nu_i}{n_1} - \frac{\mu_i}{n_2} \right)^2 \frac{n_1 n_2}{\nu_i + \mu_i} > h_\varepsilon. \qquad (6)$$

Ejemplo 1A. En el ejemplo 2.26.3 hemos descrito el mecanismo de herencia de los grupos de sangre designados por 0 (cero), A, B y AB. Dicho mecanismo es controlado por genes de tres tipos: A, B y 0. Las probabilidades de que esos genes aparezcan en una población dada se designan por p, q, $r = 1 - p - q$, respectivamente. Las probabilidades $p_i(\alpha)$, $\alpha = (p, q)$ de que una persona tenga el i-ésimo grupo de sangre se expresan a través de α según las fórmulas citadas en la tabla 1 del § 2.26.

Tenemos dos muestras X e Y con frecuencias ν_i y μ_i, $i = 1, ..., 4$ de aparición del i-ésimo grupo sanguíneo, obtenidas a consecuencia del examen de $n_1 = 353$ personas de la comunidad I, de $n_2 = 364$ personas de la comunidad II. La distribución de las personas según los grupos sanguíneos se da en la tabla 1

Tabla 1

	0	A	B	AB	Total
Comunidad I	121	120	79	33	353
Comunidad II	118	95	121	30	364
Total	239	215	200	63	717

Es necesario verificar la hipótesis de pertenencia de las comunidades examinadas a una población, o sea, la hipótesis de igualdad de las probabilidades p y q de estos grupos o, que es lo mismo, la hipótesis de igualdad de las probabilidades $p_i(\alpha)$. Este es, evidentemente, el problema de homogeneidad examinado en el ejemplo 1.

Si se verifica la coincidencia de las probabilidades de los cuatro grupos de sangre, entonces, a la estadística (veánse los capítulos precedentes)

$$\chi_1^2 = \sum_{i=1}^{4} \left(\frac{\nu_i}{n_1} - \frac{\mu_i}{n_2} \right)^2 \frac{n_1 n_2}{\nu_i + \mu_i}$$

le corresponderá la distribución χ^2 con tres grados de libertad. En nuestro caso el valor χ_1^2 constituye 11,74. El nivel realmente alcanzable (véase el § 3.4) de la desviación obtenida pasa de 0,99. Esto significa que la hipótesis de homogeneidad ha de ser rechazada desde el punto de vista del criterio $\chi_1^2 > h_{0,01}$ de nivel 0,99.

Debemos señalar que el criterio aplicado no del todo corresponde a la naturaleza del fenómeno examinado, ya que debemos verificar la coincidencia de las probabilidades p y q y no la de las probabilidades p_i de aparición de los grupos sanguíneos. Ateniéndose exactamente al teorema 1, debemos, mediante los métodos descritos en el § 2.26, calcular las e.v.m. $\hat{\alpha}_X^*$, $\hat{\alpha}_Y^*$ y $\hat{\alpha}^*$ del parámetro $\alpha = (p, q)$ con arreglo a las muestras X, Y y (X, Y), respectivamente, y utilizar la estadística

$$\chi_2^2 = 2[L(\hat{\alpha}_X^*, X) + L(\hat{\alpha}_Y^*, Y) - L(\hat{\alpha}^*, (X, Y))] =$$
$$= 2\left[\sum_{l=1}^{4} \nu_l \ln p_l(\hat{\alpha}_X^*) + \sum_{l=1}^{4} \mu_l \ln p_l(\hat{\alpha}_Y^*) - \sum_{l=1}^{4} (\nu_l + \mu_l) \ln p_l(\hat{\alpha}^*)\right]$$

que tiene, con grandes valores de n, una distribución próxima a la distribución χ^2 con dos grados de libertad. Si realizamos todos los cálculos necesarios (véase el ejemplo 2.26.3), obtendremos $\chi_2^2 \approx 11{,}04$, lo cual proporciona, para dos grados de libertad, una desviación mayor de 11,74 para tres grados de libertad.

En cuanto a la verificación de la propia hipótesis de pertenencia de X e Y a las subfamilias paramétricas $\mathbf{B}_{p(\alpha)}$, donde $p(\alpha) = p_1(\alpha), \ldots, p_4(\alpha)$, véase el ejemplo 3.17.1. Ambas muestras concuerdan bien con esta hipótesis.

Ejemplo 2. Sea $X \in \Phi_{\alpha_1, \sigma_1^2}$, $Y \in \Phi_{\alpha_2, \sigma_2^2}$, donde los puntos $\theta_l = (\alpha_l, \sigma_l^2)$ se sitúan en el entorno del punto $\theta_0 = (\alpha_0, \sigma_0^2)$. Aquí

$$I(\theta_0) = \begin{pmatrix} \sigma_0^{-2} & 0 \\ 0 & \frac{1}{2}\sigma_0^{-4} \end{pmatrix}$$

(véase el § 2.16), y examinaremos el problema de verificación de la hipótesis $H_1 = \{\theta_1 = \theta_2\}$ frente a

$$H_2^b = \left\{ \frac{(\alpha_1 - \alpha_2)^2}{\sigma_0^2} \frac{(\sigma_1^2 - \sigma_2^2)^2}{2\sigma_0^4} \geqslant \frac{b^2}{n_2} \right\}, \quad n = n_1 + n_2.$$

Tenemos $\hat{\theta}_X^* = (\overline{\mathrm{x}}, S_X^2)$, $S_X^2 = \dfrac{1}{n_1} \displaystyle\sum_{l=1}^{n_1} (\mathrm{x}_l - \overline{\mathrm{x}})^2$, $f_{\hat{\theta}_X^*}(X) = (2\pi e S_X^2)^{-n_1/2}$.

Las fórmulas análogas son válidas para la muestra Y. Seguidamente

$$\hat{\theta}^* = (\overline{\mathrm{z}}, S_{X,Y}^2), \ \overline{\mathrm{z}} = \frac{\left(\sum_{l=1}^{n_1} \mathrm{x}_l + \sum_{l=1}^{n_2} \mathrm{y}_l\right)}{n_1 + n_2} = a\overline{\mathrm{x}} + (1 - a)\overline{\mathrm{y}}, \tag{7}$$

$$S_{X,Y}^2 = \frac{1}{n_1 + n_2}\left[\sum_{l=1}^{n_1} (\mathrm{x}_l - \overline{\mathrm{z}})^2 + \sum_{l=1}^{n_2} (\mathrm{y}_l - \overline{\mathrm{z}})^2\right] =$$
$$= aS_X^2 = (1 - a)S_Y^2 + (1 - a)a(\overline{\mathrm{x}} - \overline{\mathrm{y}})^2,$$

donde $a = n_1/(n_1 + n_2)$, $f_{\hat{\beta}^*}(X) f_{\hat{\beta}^*}(Y) = (2\pi e S^2_{X,Y})^{-\frac{1}{2}(n_1+n_2)}$. Ahora bien, para verificar H_1 frente a H_2^b, como criterio asintóticamente minimax utilizaremos el criterio

$$\frac{S^2_{X,Y}}{S_X^{2a} S_Y^{2(1-a)}} > e^{h_\varepsilon/(n_1+n_2)},$$

donde h_ε es la cuantila de la distribución χ^2 con dos grados de libertad. Le proponemos al lector que halle, en calidad de ejercicio, el criterio asintóticamente equivalente que tiene la forma (5).

3. Criterios asintóticamente minimax para el problema de homogeneidad al existir un parámetro obstaculizador. En éste y en los apartados posteriores supondremos, para abreviar, que los volúmenes de las muestras X e Y coinciden: $n_1 = n_2$. Esta limitación no tiene importancia. En el caso de $n_1/n_2 = r_1/r_2$ (r_1 y r_2 son enteros) el lector puede liberarse por sí mismo de esta limitación así como se hizo en la observación 1 del teorema 1.

Así pues, supongamos que se dan dos muestras $X \in \mathbf{P}_{\theta_1}$ e $Y \in \mathbf{P}_{\theta_2}$, $\theta_i = (u_i, v_i)$, $i = 1, 2$, de volúmenes $n_1 = n_2 = n$. Se verifica la hipótesis $\{u_1 = u_2\}$ frente a $\{u_1 \neq u_2\}$ suponiendo que conocemos $v_1 = v_2 = v$ y v. La dimensión u_i se designa por l, $l < k$.

Introduzcamos un nuevo parámetro $\overline{\theta} = (u_1, u_2, v)$. Representemos la muestra unida (X, Y) como una muestra de volumen n con observaciones $(\mathrm{x}_1, \mathrm{y}_1), \ldots, (\mathrm{x}_n, \mathrm{y}_n)$ cuya densidad de distribución es igual a $f_{\overline{\theta}}(x, y) = f_{(u_1, v)}(x) f_{(u_2, v)}(y)$. Para esta familia paramétrica, el problema sometido a investigación equivale al problema de verificación de la hipótesis H_1, que consiste en el hecho de que el valor de $\overline{\theta}$ se encuentra en la "curva" $\overline{\theta} = g(\theta_1) = (u_1, u_1, v)$ frente a la alternativa adicional. La matriz $G = \left\| \dfrac{\partial g_i}{\partial \theta_{1j}} \right\|$, $i = 1, \ldots, k+1$, $j = 1, \ldots, k$, tiene la forma $\begin{pmatrix} E_l & 0 \\ & E_k \end{pmatrix}$, donde arriba se halla la matriz unidad de orden 1, y abajo, la matriz unidad de orden k, así que el rango de G es igual a k.

Al igual que en el apartado anterior, consideraremos que el parámetro θ ha sido localizado cerca del punto $\theta_0 = (u_0, v_0)$. Introduzcamos el parámetro $\tau = \tau(\overline{\theta}) = (\tau', \tau'', \tau''') = (u_1 - u_0, u_2 - u_1, v - v_0)$. La aplicación inversa $\overline{\theta} = \overline{\theta}(\tau)$ siempre existe y sus coordenadas son $u_1 = \tau' + u_0$, $u_2 = \tau'' + \tau' + u_0$, $v = \tau''' + v_0$. Pongamos $\tau = \gamma/\sqrt{n}$, $\gamma = (\gamma', \gamma'', \gamma''')$.

Para el nuevo parámetro τ(o γ), la hipótesis de homogeneidad tiene la forma $H_1 = \{\gamma'' = 0\}$. En calidad de alternativa examinemos la hipótesis "aislada" $H_2^b = \{\gamma'' I_1(\theta_0) \gamma^T \geqslant b^2\}$, donde $I_1(\theta)$ es la submatriz de la matriz inicial de información de Fisher $I(\theta)$, formada por sus primeras l filas y columnas.

Teorema 2. *Supongamos que en el entorno del punto θ_0, la familia $\{\mathbf{P}_\theta\}$ satisface las condiciones (RR). Entonces, el criterio de relación de verosi-*

militud

$$R_1(X, Y) = \frac{\sup\limits_{(u_1, u_2, v)} f_{(u_1, v)}(X) f_{(u_2, v)}(Y)}{\sup\limits_{\theta} f_\theta(X) f_\theta(Y)} > e^{h_\varepsilon/2} \tag{8}$$

es el criterio asintóticamente minimax, de nivel asintótico $1 - \varepsilon$, *para verificar* $H_1 = \{u_1 = u_2\}$ *frente a*

$$H_2^b = \{(u_1 - u_2) I_1(\theta_0)(u_1 - u_2)^T \geqslant b^2/n\}, \tag{9}$$

con un valor común de $v_1 = v_2 = v$ *y con cualquier* $b > 0$. *Aquí* h_ε *es una cuantila del orden de* $1 - \varepsilon$ *de la distribución* χ^2 *con* l *grados de libertad. (Tal será la distribución límite* $2 \ln R_1(X, Y)$ *en la hipótesis* H_1).

Designemos por $\overline{\theta}^*$ *el valor del parámetro* $\overline{\theta}$ *con el que se alcanza el valor máximo del numerador* en (8), y por $\theta^* = (u^*, v^*)$, *el valor de* θ *con el que se alcanza el valor máximo del denominador. Representemos la matriz* $I(\theta)$ *en la forma*

$$I(\theta) = \begin{pmatrix} I_1(\theta) & I_{21}(\theta) \\ I_{12}(\theta) & I_{22}(\theta) \end{pmatrix}.$$

Entonces, el criterio

$$(\overline{\theta}^* - (u^*, u^*, v^*)) \overline{I}(\overline{\theta}^*)(\overline{\theta}^* - (u^*, u^*, v^*))^T > h_\varepsilon/n, \tag{10}$$

donde

$$\overline{I}(\overline{\theta}) = \begin{pmatrix} I_1(\theta_1) & 0 & I_{21}(\theta_1) \\ 0 & I_1(\theta_2) & I_{21}(\theta_2) \\ I_{12}(\theta_1) & I_{12}(\theta_2) & I_{22}(\theta_1) + I_{22}(\theta_2) \end{pmatrix}, \tag{11}$$

será asintóticamente equivalente a (8).

Demostración. Este teorema también es el corolario directo del teorema 3.15.4. Sólo queda aclarar la estructura de la matriz $\overline{I}(\overline{\theta})$ para la muestra (X, Y) del parámetro "unido" $\overline{\theta}$ y de la matriz $\mathbf{M}_2$. Tenemos

$$l \equiv \ln f_{\overline{\theta}}(x, y) = l(x, (u_1, v)) + l(y, (u_2, v)).$$

Designemos por t_i, $i = 1, \dots, k + l$, las coordenadas del vector $\overline{\theta}$. Entonces

$$\frac{\partial l}{\partial t_i} = \begin{cases} \dfrac{\partial l(x, (u_1, v))}{\partial t_i}, & 0 < i \leqslant l, \\ \dfrac{\partial l(y, (u_2, v))}{\partial t_i}, & l < i \leqslant 2l, \\ \dfrac{\partial l(x, (u_1, v))}{\partial t_i} + \dfrac{\partial l(y, (u_2, v))}{\partial t_i}, & 2l < i \leqslant k + l; \end{cases}$$

de aquí se obtiene (11) sin dificultad.

La matriz $\mathbf{M}_2$ para la familia paramétrica $\mathbf{P}_{\bar{\theta}(0,\ \beta,\ 0)} = \mathbf{P}_{(u_0,\ u_0+\beta,\ v_0)}$ en el punto $\beta = 0$ se calcula análogamente. La misma es igual a $I_1(\theta_0)$ y corresponde a la submatriz media de la matriz $\bar{I}(\bar{\theta}_0)$. ◁

En los ejemplos expuestos consideraremos que los volúmenes de las muestras n_1 y n_2 son arbitrarios.

Ejemplo 3. Sea $X \in \Phi_{\alpha_1,\ \sigma^2}$, $Y \in \Phi_{\alpha_2,\ \sigma^2}$. Es necesario verificar la hipótesis $H_1 = \{\alpha_1 = \alpha_2\}$ cuando se desconoce σ^2. Para determinar los criterios asintóticamente minimax con ayuda del teorema 2 necesitamos hallar la estadística $R_1(X,\ Y)$ en (8), donde en nuestro caso $u_1 = \alpha_i$, $v = \sigma^2$, $\bar{\theta} = (\alpha_1,\ \alpha_2,\ \sigma^2)$. Tenemos $\ln f_{(\alpha_1,\ \sigma^2)}(X) f_{(\alpha_2,\ \sigma^2)}(Y) = -\dfrac{1}{2}(n_1 + n_2)\ln(2\pi\sigma^2) - \dfrac{1}{2\sigma^2}\sum_{i=1}^{n_1}(x_i - \alpha_1)^2 - \dfrac{1}{2\sigma^2}\sum_{i=1}^{n_2}(y_i - \alpha_2)^2$. Reduciendo a cero las derivadas de esta función respecto a α_1, α_2 y σ^2, y resolviendo las ecuaciones obtenidas, hallamos según las designaciones del ejemplo 2)

$$\bar{\theta}^* = (\bar{x},\ \bar{y},\ aS_X^2 + (1 - a)S_Y^2), \quad a = \frac{n_1}{n_1 + n_2}, \tag{12}$$

$$f_{\bar{\theta}^*}(X,\ Y) = [2\pi e(aS_X^2 + (1 - a)S_Y^2)]^{-(n_1+n_2)/2}.$$

Procediendo del mismo modo con la función $\ln f_\theta(X) f_\theta(Y) = \ln f_{(\alpha,\sigma^2)}(X) f_{(\alpha,\sigma^2)}(Y)$, obtenemos (véase el ejemplo 2)

$$\theta^* = (\bar{z},\ S_{X,Y}^2),$$

$$f_{\theta^*}(X) f_{\theta^*}(Y) = (2\pi e S_{X,Y}^2)^{-\frac{1}{2}(n_1+n_2)}. \tag{13}$$

Ahora bien, el criterio asintóticamente óptimo tiene la forma

$$\frac{S_{X,Y}^2}{aS_X^2 + (1 - a)S_Y^2} > e^{h_\varepsilon/(n_1+n_2)}$$

o bien (veáse (7))

$$\frac{\sqrt{a(1 - a)}\,|\bar{x} - \bar{y}|}{\sqrt{aS_X^2 + (1 - a)S_Y^2}} > \sqrt{\frac{h_\varepsilon}{n_1 + n_2}},$$

donde h_ε es una cuantila del orden de $1 - \varepsilon$ de la distribución χ^2 con un solo grado de libertad, así que $\sqrt{h_\varepsilon}$ se puede sustituir por el valor de $\lambda_{\varepsilon/2}$ para el cual $\Phi_{0,\ 1}(-\lambda_{\varepsilon/2},\ \lambda_{\varepsilon/2}) = 1 - \varepsilon$. Es fácil notar que el primer miembro de la desigualdad

$$\frac{\sqrt{a(1 - a)(n_1 + n_2)}\,|\bar{x} - \bar{y}|}{\sqrt{aS_X^2 + (1 - a)S_Y^2}} > \lambda_{\varepsilon/2} \tag{14}$$

que define el criterio asintóticamente minimax, después de sustituir $|\bar{x} - \bar{y}|$ por $\bar{x} - \bar{y}$ será asintóticamente normal con los parámetros (0, 1) de una variable aleatoria.

Pero este criterio puede ser exacto (o sea, puede tener con exactitud un nivel dado de antemano). Efectivamente, en virtud de los resultados del § 2.32, en el caso de la hipótesis H_1,

$$\sqrt{\frac{n_1 n_2}{n_1 + n_2}}\,\frac{\bar{x} - \bar{y}}{\sigma} \in \Phi_{0,1},$$

$$\frac{(n_1 + n_2) a S_X^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n_1} (x_i - \bar{x})^2 \in \mathbf{H}_{n_1 - 1},$$

$$\frac{(n_1 + n_2)(1 - a) S_Y^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n_2} (y_i - \bar{y})^2 \in \mathbf{H}_{n_2 - 1}.$$

En vista de que las tres variables aleatorias son independientes, la relación

$$(\bar{x} - \bar{y}) \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left[\frac{n_1 + n_2}{n_1 + n_2 - 2} (a S_X^2 + (1 - a) S_Y^2)\right]^{-1/2} =$$

$$= \frac{(\bar{x} - \bar{y})\sqrt{a(1 - a)(n_1 + n_2 - 2)}}{\sqrt{a S_X^2 + (1 - a) S_Y^2}} \in \mathbf{T}_{n_1 + n_2 - 2}$$

tiene distribución de Student con $n_1 + n_2 - 2$ grados de libertad. Así pues, el criterio (compárese con (14))

$$\frac{(\bar{x} - \bar{y})\sqrt{a(1 - a)(n_1 + n_2 - 2)}}{\sqrt{a S_X^2 + (1 - a) S_Y^2}} > \tau_\varepsilon,$$

donde τ_ε es tal, que $\mathbf{T}_{n_1 + n_2 - 2}(-\tau_\varepsilon, \tau_\varepsilon) = 1 - \varepsilon$ tendrá un nivel de significación exactamente igual a $1 - \varepsilon$ y el mismo podrá ser utilizado para cualesquiera valores (y no sólo grandes) de n_1, n_2. Este criterio, que se denomina criterio de Student, también posee ciertas propiedades de optimización exacta (y no sólo asintótica) (veáse [57]).

Ejemplo 4. Sea $X \in \Phi_{\alpha, \sigma_1^2}$, $Y \in \Phi_{\alpha, \sigma_2^2}$. La hipótesis $\{\sigma_1 = \sigma_2\}$ se verifica cuando se desconoce α. Procediendo del mismo modo que en el ejercicio anterior, llegaremos al valor R_1 en (8), cuyo denominador equivale al del ejemplo anterior, y el numerador es igual a

$$\sup_{\alpha, \sigma_1 \sigma_2} f_{(\alpha, \sigma_1^2)}(X) f_{(\alpha, \sigma_2^2)}(Y). \tag{15}$$

Escribiendo las ecuaciones para el punto del valor máximo, obtenemos

$$\sigma_1^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(x_i-\alpha)^2 = S_X^2 + (\bar{x}-\alpha)^2,$$

$$\sigma_2^2 = \frac{1}{n_2}\sum_{i=1}^{n_2}(y_i-\alpha)^2 = S_Y^2 + (\bar{y}-\alpha)^2,$$

$$\frac{n_1}{\sigma_1^2}(\bar{x}-\alpha) + \frac{n_2}{\sigma_2^2}(\bar{y}-\alpha) = 0.$$

De aquí, poniendo

$$p = \frac{a}{\sigma_1^2}\cdot\frac{1}{a/\sigma_1^2 + (1-a)/\sigma_2^2} \in (0,\ 1), \tag{16}$$

hallamos

$$\alpha = p\bar{x} + (1-p)\bar{y},$$
$$\sigma_1^2 = S_X^2 + (1-p)^2\Delta^2, \quad \sigma_2^2 = S_Y^2 + p^2\Delta^2,$$

donde, para abreviar, hemos supuesto que $\Delta = \bar{x} - \bar{y}$; p puede considerarse como la solución de la ecuación (16) o

$$p = \frac{a(S_Y^2 + p^2\Delta^2)}{a(S_Y^2 + p^2\Delta^2) + (1-a)(S_X^2 + (1-p)^2\Delta^2)}.$$

Como el máximo en (15) es igual a

$$(2\pi e)^{-(n_1+n_2)/2}(S_X^2 + (1-p)^2\Delta^2)^{-n_1/2}(S_Y^2 + p^2\Delta^2)^{-n_2/2}, \tag{17}$$

comparándolo con (13) y (7), obtenemos el criterio asintóticamente minimax

$$\frac{aS_X^2 + (1-a)S_Y^2 + a(1-a)\Delta^2}{(S_X^2 + (1-p)^2\Delta^2)^2(S_Y^2 + p^2\Delta^2)^{1-a}} > e^{h_\varepsilon(n_1+n_2)} \tag{18}$$

o bien

$$\frac{aS_X^2 + (1-a)S_Y^2}{S_X^{2a}S_Y^2(1-a)} > e^{h_\varepsilon/(n_1+n_2)}A^{-1}, \tag{19}$$

donde $A = \dfrac{1 + \dfrac{a(1-a)\Delta^2}{aS_X^2 + (1-a)S_Y^2}}{(1 + (1-p)^2\Delta^2/S_X^2)^a(1 + p^2\Delta^2/S_Y^2)^{1-a}}$, h_ε es una cuantila de la distribución χ^2 con un solo grado de libertad. Aquí

$$\Delta^2 = (\sigma_1^2/n_1 + \sigma_2^2(n_2)\xi^2,\ \xi \in \Phi_{0,1},\ S_X^2/\sigma_1^2 \underset{P}{\to} 1,\ S_Y^2/\sigma_2^2 \underset{P}{\to} 1,\ \sigma_1^2/\sigma_2^2 \to 1,$$

$p \to a$ (para abreviar podemos considerar que $a = \dfrac{n_1}{n_1 + n_2}$ es fijo), $\ln A \underset{P}{\to} 0$ para cada una de las hipótesis semejantes que se examinan. Por consiguiente, el segundo miembro en (19) tiene la forma

$$1 + \frac{h_\varepsilon + \delta_n}{n_1 + n_2}, \qquad \delta_n \underset{P}{\to} 0.$$

El primer miembro de (19) es la relación entre la media aritmética y la media geométrica de los valores de S_X^2 y S_Y^2. Si se designa $S_X^2/S_Y^2 = Z^2$, la desigualdad inversa a (19) puede ser escrita en la forma

$$\frac{aZ^2 + (1-a)}{Z^{2a}} - 1 \leqslant \frac{h_\varepsilon + \delta_n}{n_1 + n_2}. \tag{20}$$

Aquí, en el primer miembro se halla la función de Z convexa hacia abajo (para evidenciar la exposición podemos considerar $a \leqslant 1/2$) que tiene un cero múltiplo en el punto $Z = 1$. Como el segundo miembro de esta desigualdad es pequeño, conviene hallar la solución en forma de $Z^2 = 1 + \zeta$ cuando ζ es pequeño. Utilizando el desarrollo en serie respecto a las potencias de ζ, y eliminando los términos del tercero y mayores órdenes de pequeñez, obtenemos, para las fronteras ζ_1, ζ_2 del intervalo donde es válida (20), los valores

$$\zeta_1 = \sqrt{\frac{2(h_\varepsilon + \delta_n')}{a(1-a)(n_1+n_2)}}, \quad \zeta_2 = \sqrt{\frac{2(h_\varepsilon + \delta_n')}{a(1-a)(n_1+n_2)}},$$

$$\delta_n' \underset{P}{\to} 0, \quad \delta_n'' \underset{P}{\to} 0.$$

Esto significa que, si volvemos a las variables iniciales, el dominio

$$\sqrt{\frac{1}{2}\, a(1-a)(n_1+n_2)}\; |S_X^2/S_Y^2 - 1| > \sqrt{h_\varepsilon} = \lambda_{\varepsilon/2} \tag{21}$$

(λ_ε ha sido definido en el ejemplo 3) definirá el criterio asintóticamente equivalente a (18) y, por lo tanto, asintóticamente minimax.

Aquí al igual que en el ejemplo 3, podemos hacer que el criterio obtenido sea exacto, ya que conocemos la distribución precisa de la estadística S_X^2/S_Y^2. En efecto,

$$n_1 S_X^2/\sigma_1^2 \in \mathbf{H}_{n_1-1}, \; n_2 S_Y^2/\sigma_2^2$$

y en el caso de la hipótesis $H_1 = \{\sigma_1 = \sigma_2\}$,

$$\frac{n_1 S_X^2}{n_2 S_Y^2} \in \mathbf{F}_{n_1-1,\, n_2-1},$$

donde $\mathbf{F}_{n_1-1,n_2-1}$ es la distribución de Fisher introducida en el § 2.2 y tabulada en los manuales de estadística matemática. Esto significa que es posible calcular el nivel exacto de significación del criterio (21) y aplicarlo para cualesquiera n_1 y n_2 (las propiedades exactas de optimización de este criterio se exponen en [57]). Si son grandes los valores de n_1 y n_2, el primer miembro en (21) (sin signo de valor absoluto) es asintóticamente normal con parámetros (0, 1).

4. Criterio asintóticamente minimax para el problema de homogeneidad parcial. Supongamos que $X \in \mathbf{P}_{\theta_1}$, $Y \in \mathbf{P}_{\theta_2}$, $\theta_i = (u_i, v_i)$, $i = 1, 2$. Se verifica la hipótesis $\{u_1 = u_2\}$ frente a $\{u_1 \neq u_2\}$ cuando los valores de v_1 y v_2 en las muestras X e Y pueden ser cualesquiera. La dimensión u_i, al igual que antes, se designa por l, $l > k$.

Introduzcamos el nuevo parámetro $\overline{\theta} = (\theta_1, \theta_2) = (u_1, v_1, u_2, v_2)$ de dimensión $2k$. Al igual que antes, representemos la muestra (X, Y) (cuando $n_1 = n_2 = n$) como muestra con observaciones $(\mathrm{x}_1, \mathrm{y}_1)$, ..., $(\mathrm{x}_n, \mathrm{y}_n)$ de densidad

$$f_{\overline{\theta}}(x, y) = f_{(u_1, v_1)}(x) f_{(u_2, v_2)}(y).$$

Para esta familia, el problema de homogeneidad parcial equivale al problema de verificación de la hipótesis H_1, el cual consiste en que $\overline{\theta}$ permanece en la "curva" $\theta = g(\alpha) = (u_1, v_1, u_1, v_2)$, donde $\alpha = (u_1, v_1, v_2)$ es el "subparámetro" de dimensión $2k - l$. Le proponemos al lector que escriba, siguiendo los razonamientos de los dos apartados anteriores, la matriz $G = \left\| \dfrac{\partial g_i}{\partial \alpha_j} \right\|$, $i = 1, ..., 2k$, $j = 1, ..., 2k - l$. Su rango es igual a $2k - l$.

Al igual que en los apartados 2 y 3, consideraremos "localizado" el problema cerca del punto $\theta_0 = (u_0, v_0)$. A la par con θ introduzcamos el parámetro $\tau = \tau(\overline{\theta}) = (\tau', \tau'', \tau''', \tau^{\mathrm{IV}}) = (u_1 - u_0, v_1 - v_0, u_2 - u_1, v_2 - v_0)$. La transformación inversa $\overline{\theta} = \overline{\theta}_{(\tau)}$ tiene las coordenadas

$$u_1 = \tau' + u_0,\ v_1 = \tau'' + v_0,\ u_2 = \tau''' + u_0,$$
$$v_2 = \tau^{\mathrm{IV}} + v_0.$$

Si se pone $\tau = \gamma/\sqrt{n}$, $\gamma = (\gamma', \gamma'', \gamma''', \gamma^{\mathrm{IV}})$, la hipótesis H_1 tendrá la forma $H_1 = \{\gamma''' = 0\}$. En calidad de alternativa consideraremos la hipótesis "aislada" $H_2^b = \{\gamma''' I_1(\theta_0) \gamma'''^T \geqslant b^2\}$, donde $I_1(\theta)$ tiene el mismo sentido que en el teorema 2.

Teorema 3. *Supongamos que en el entorno del punto θ_0, la familia $\{\mathbf{P}_\theta\}$ satisface las condiciones (RR). Entonces, el criterio de la relación de verosimilitud*

$$R_1(X, Y) = \frac{\sup\limits_{(\theta_1, \theta_2)} f_{\theta_1}(X) f_{\theta_2}(Y)}{\sup\limits_{(u, v_1, v_2)} f_{(u, v_1)}(X) f_{(u, v_2)}(Y)} > e^{h_\varepsilon/2} \tag{22}$$

es el criterio asintóticamente minimax de nivel asintótico $1 - \varepsilon$ *para verificar* H_1 *frente a la hipótesis* H_2^b *definida en* (9), *para los valores arbitrarios de* v_1 *y* v_2. *El valor de* h_ε *aquí es el mismo que en el teorema* 2.

La demostración de este teorema repite los razonamientos de los apartados precedentes y asimismo se basa por completo en el teorema 3.15.4. Le dejamos al lector que él mismo determine la matriz de información de Fisher $\overline{I}(\overline{\theta})$ para el parámetro $\overline{\theta}$, y la matriz $\mathbf{M}_2$ para la familia de densidad $f_{\overline{\theta}(0,\,0,\,\beta,\,0)} = f_{(u_0,\,v_0,\,u_0+\beta,\,v_0)}$ en el punto $\beta = 0$.

Con ayuda de la matriz $\overline{I}(\hat{\theta}_X^*,\ \hat{\theta}_Y^*))$ y los vectores $(\hat{\theta}_X^*,\ \hat{\theta}_Y^*) - (u^*,\ v_1^*,\ u^*,\ v_2^*)$, donde $(\hat{\theta}_X^*,\ \hat{\theta}_Y^*)$ y $(u^*,\ v_1^*,\ v_2^*)$ son los vectores en los que se alcanzan los valores máximos del numerador y el denominador en (22), es posible, como antes, mediante el teorema 3.15.4 (véase (3.15.12), construir el criterio asintóticamente equivalente que utiliza la forma cuadrática de las estimaciones introducidas. ◁

Ejemplo 5. *Comparación de las varianzas de las poblaciones normales. Sea* $X \in \Phi_{\alpha_1,\ \sigma_1^2}$, $Y \in \Phi_{(\alpha_2,\ \sigma_2^2)}$, $H_1 = \{\sigma_1 = \sigma_2\}$. Aquí, los cálculos son mucho más fáciles que en el ejemplo 4, ya que conocemos el valor del numerador en (22) (al igual que el vector $(\hat{\theta}_X^*,\ \hat{\theta}_Y^*) = (\overline{x},\ S_X^2,\ \overline{y},\ S_Y^2))$, y el valor del denominador ha sido hallado en el ejemplo 3 (véase (12)). La desigualdad (22) aquí tendrá la forma

$$\frac{aS_X^2 + (1-a)S_Y^2}{S_X^{2a}S_Y^{2(1-a)}} > e^{h_\varepsilon/(n_1+n_2)}.$$

Comparando esto con (19) y con los planteamientos posteriores, llegaremos a los mismos criterios y a las mismas deducciones que en el ejemplo 4.

Ejemplo 6. *Problema de Behrens — Fisher acerca de la comparación de las medias de dos poblaciones normales.* Sea $X \in \Phi_{\alpha_1,\ \sigma_1^2}$, $Y \in \Phi_{\alpha_2,\ \sigma_2^2}$, $H_1 = \{\alpha_1 = \alpha_2\}$ y supongamos que los valores σ_1 y σ_2 son arbitrarios. Para este ejemplo, el numerador en (22) es el mismo que en el párrafo anterior, y el denominador fue hallado en el ejemplo 4 (véase (17); allí éste era el numerador para (8)).

Por consiguiente, el criterio asintóticamente minimax tiene la forma

$$\left(\frac{S_X^2 + (1-p)^2\Delta^2}{S_X^2}\right)^a \left(\frac{S_Y^2 + p^2\Delta^2}{S_Y^2}\right)^{1-a} > e^{h_\varepsilon/(n_1+n_2)}; \tag{23}$$

aquí $\Delta = \overline{x} - \overline{y}$ es representable en la forma

$$\Delta = (\alpha_1 - \alpha_2) + \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\,\xi, \quad \xi \in \Phi_{0,1}.$$

$$S_X^2/\sigma_1^2 \underset{p}{\to} 1, \quad S_Y^2/\sigma_2^2 \underset{p}{\to} 1,$$

así que $\Delta \xrightarrow[p]{} 0$ para la hipótesis H_1. Esta relación, evidentemente, también conserva su validez para cada una de las alternativas semejantes. Para hallar un criterio más simple en cuanto a su forma y que equivalga asintóticamente a (23), en ambos miembros de la desigualdad (23) separaremos sus partes principales. Obtendremos

$$\frac{a(1-p)^2\Delta^2}{S_X^2} + \frac{(1-a)p^2\Delta^2}{S_Y^2} + \Delta^4 \varrho_n > \frac{h_\varepsilon}{n_1+n_2} + O\left(\frac{1}{(n_1+n_2)^2}\right),$$

donde $\varrho_n \xrightarrow[p]{} \varrho = \text{const}$. Teniendo en cuenta que

$$p = \frac{aS_Y^2}{aS_Y^2 + (1-a)S_X^2} + \Delta^2 \varrho_n',$$

$\varrho_n' \xrightarrow[p]{} \varrho' = \text{const}$, obtenemos

$$\frac{a(1-a)^2 S_X^2 \Delta^2 (n_1+n_2) + a^2(1-a) S_Y^2 \Delta^2 (n_1+n_2)}{(aS_Y^2 + (1-a)S_X^2)^2} +$$

$$+ \Delta^4 (n_1+n_2) \varrho_n'' > h_\varepsilon + O\left(\frac{1}{n_1+n_2}\right),$$

donde $\varrho_n'' \to \varrho'' = \text{const}$, $\Delta^4(n_1+n_2) \xrightarrow[p]{} 0$. Equivalentemente esto se puede escribir de la forma siguiente:

$$\frac{\Delta^2(n_1+n_2)}{S_X^2/a + S_Y^2/(1-a)} > h_\varepsilon + \delta_n, \quad \delta_n \xrightarrow[p]{} 0.$$

De aquí se deduce que el criterio

$$\frac{|\bar{x} - \bar{y}|\sqrt{n_1+n_2}}{\sqrt{S_X^2/a + S_Y^2/(1-a)}} > \sqrt{h_\varepsilon} = \lambda_{\varepsilon/2} \tag{24}$$

es asintóticamente equivalente a (23) y, por lo tanto, asintóticamente minimax para el problema de Behrens — Fisher. Aquí $\lambda_{\varepsilon/2}$ tiene el mismo sentido que en el ejemplo 4. A distinción de los ejemplos 2—4, aquí la distribución antelímite de la estadística en el primer miembro (24) depende, para la hipótesis H_1, de los parámetros σ_1^2 y σ_2^2.

5. Algunos otros problemas. Aquí señalaremos dos clases más de problemas cuya solución asintótica puede ser hallada con ayuda del teorema 3.15.4.

1) A la primera clase de problemas pertenecen aquéllos que generalizan los problemas de los apartados 2—4 para el caso cuando se verifican las hipótesis de tipo $\{\theta_1 = f(\theta_2)\}$ (por ejemplo, $\{\theta_1 = a + b\theta_2\}$) en condiciones del apartado 2, y de tipo $\{u_1 = f(u_2)\}$ en condiciones de los apartados 3 y 4. Es fácil notar que los planteamientos de los apartados 2—4 se extienden a este caso más general.

2) A la segunda clase de problemas pertenecen aquéllos que constan de tres muestras y más. Examinemos, por ejemplo, el problema de homogeneidad para tres muestras. Supongamos que $X \in \mathbf{P}_{\theta_1}$, $Y \in \mathbf{P}_{\theta_2}$, $Z \in \mathbf{P}_{\theta_3}$. Se verifica la hipótesis $H_1 = \{\theta_1 = \theta_2 = \theta_3\}$ frente a la alternativa adicional. Supongamos, para abreviar, que los volúmenes n_1, n_2 y n_3 de las muestras son iguales a $n_1 = n_2 = n_3 = n$. Examinemos la muestra unida (X, Y, Z) como una muestra de volumen n con observaciones (x_1, y_1, z_1), ..., (x_n, y_n, z_n) de densidad $f_{\overline{\theta}}(x, y, z) = f_{\theta_1}(x)f_{\theta_2}(y)f_{\theta_3}(z)$, donde $\overline{\theta} = (\theta_1, \theta_2, \theta_3)$. Entonces, la hipótesis H_1 será equivalente al hecho de que $\overline{\theta}$ permanece en la "curva" $\overline{\theta} = g(\alpha)$, $\alpha \equiv \theta_1$, $g(\alpha) = (\alpha, \alpha, \alpha)$. Vemos que el problema de nuevo se reduce al problema examinado en el teorema 3.15.4.

§ 2. Problema de homogeneidad en el caso general

1. Planteamiento del problema. En este párrafo examinaremos dos muestras X e Y de volúmenes n_1 y n_2, respectivamente, sin suponer que las mismas pertenecen a cualquier familia paramétrica.

El problema de homogeneidad de las muestras X e Y, en el caso general consiste en lo siguiente. Designemos por $\mathbf{P}_1$ y $\mathbf{P}_2$ las distribuciones de las muestras X e Y: $X \in \mathbf{P}_1$, $Y \in \mathbf{P}_2$. Se verifica la hipótesis $H_1 = \{\mathbf{P}_1 = \mathbf{P}_2\}$ frente a $H_2 = \{\mathbf{P}_1 \neq \mathbf{P}_2\}$. Evidentemente, ambas hipótesis son compuestas. Las distribuciones $\mathbf{P}_1$ y $\mathbf{P}_2$ pueden elegirse de una familia dada $\mathscr{P}$ o ser arbitrarias. El principio general de construcción del criterio estadístico para verificar H_1 frente a H_2 es el mismo que en el capítulo 3. Al igual que en el § 1, la diferencia sólo consiste en que aquí este principio se basa en la muestra unida (X, Y), así que $\pi = \pi(X, Y)$ es la probabilidad de aceptar H_2 para una muestra dada (X, Y). En el caso no randomizado ($\pi = 0$ ó 1), el criterio π es definido por una región crítica $\Omega \subset \mathscr{X}^{n_1+n_2}$ tal, que para $(X, Y) \in \Omega$ se acepta H_2. El número

$$1 - \varepsilon = \inf_{\mathbf{P}_1 \in \mathscr{P}} \mathbf{P}_1 \times \mathbf{P}_1(X, Y) \notin \Omega)$$

se llama *nivel de significación*, y el valor

$$\beta_\pi(\mathbf{P}_1, \mathbf{P}_2) = \mathbf{P}_1 \times \mathbf{P}_2(X, Y) \in \Omega), \quad \mathbf{P}_1 \in \mathscr{P}, \quad \mathbf{P}_2 \in \mathscr{P},$$

se denomina potencia del criterio π en el "punto" $(\mathbf{P}_1, \mathbf{P}_2)$.

El criterio π se denomina criterio *conciliable* si $\beta_n(\mathbf{P}_1, \mathbf{P}_2) \to 1$ cuando $n_1 \to \infty$, $n_2 \to \infty$ y para todas $\mathbf{P}_1 \neq \mathbf{P}_2$, $\mathbf{P}_1 \in \mathscr{P}$, $\mathbf{P}_2 \in \mathscr{P}$.

Ya sabemos que con el crecimiento de n_1 y n_2, las distribuciones empíricas $\mathbf{P}_X^*$, $\mathbf{P}_Y^*$, correspondientes a las muestras X e Y, se aproximan indefinidamente a $\mathbf{P}_1$ y $\mathbf{P}_2$, respectivamente. Por eso, la base natural para construir los criterios de homogeneidad es el uso de distintos tipos de "distancias" $d(\mathbf{P}_X^*, \mathbf{P}_Y^*)$ entre $\mathbf{P}_X^*$ y $\mathbf{P}_Y^*$, donde d satisface las mismas condiciones generales que hemos descrito en el § 3.12. En este caso revisten interés especial los criterios no paramétricos y asintóticamente no paramétricos que se definen del modo siguiente.

Sea $d(\mathbf{P}, \mathbf{Q})$ cierta distancia (no obligatoriamente métrica) en el espacio de distribuciones. Si la probabilidad

$$\mathbf{P}_1 \times \mathbf{P}_1(d(\mathbf{P}_X^*, \mathbf{P}_Y^*) > c) = \varepsilon \tag{1}$$

no depende de la muestra $\mathbf{P}_1$, entonces el criterio π, definido por las igualdades

$$\pi(X, Y) = \begin{cases} 0, \text{ si } d(\mathbf{P}_X^*, \mathbf{P}_Y^*) \leqslant c, \\ 1, \text{ en el caso contrario,} \end{cases} \tag{2}$$

se llama criterio *no paramétrico.* Es evidente que el criterio no paramétrico construido tendrá un igual a nivel $1 - \varepsilon$.

Así mismo se determinan los criterios no paramétricos cuando (1) se conserva asintóticamente al introducir la operación $\lim\limits_{n_1 \to \infty,\ n_2 \to \infty}$ en el primer miembro. En este caso el criterio (2) tendrá un nivel asintótico igual $1 - \varepsilon$. Cuando falta la no parametricidad (exacta o asintótica) es muy difícil construir los criterios de verificación de la homogeneidad de un nivel dado.

Examinemos algunos criterios principales de verificación de la homogeneidad.

2. Criterio de Kolmogórov — Smirnov. Supongamos que $\mathbf{P}_1$ y $\mathbf{P}_2$ pertenecen a la clase $\mathscr{P}$ de todas las distribuciones continuas en una recta, y que F_X^* y F_Y^* son funciones empíricas de distribución, correspondientes a $\mathbf{P}_X^*$ y $\mathbf{P}_Y^*$. En calidad de distancia $d(\mathbf{P}_X^*, \mathbf{P}_Y^*)$, el criterio de Kolmogórov — Smirnov considera la estadística

$$D_{n_1, n_2} = \sup_t |F_X^*(t) - F_Y^*(t)|.$$

El criterio $D_{n_1, n_2} > c$, construido con ayuda de la estadística D_{n_1, n_2} no es paramétrico. En efecto, supongamos que es cierta la hipótesis H_1 y que $F(t)$ es la función general de distribución de X e Y. La estadística D_{n_1, n_2} se puede escribir de la forma siguiente:

$$D_{n_1 n_2} = \sup_t |G_X^*(F(t)) - G_Y^*(F(t))|, \tag{3}$$

donde $G_X^*(u) = F_X^*(F^{-1}(u))$ es la función empírica de distribución que corresponde a la distribución uniforme en [0, 1] (veánse los §§ 1.6 y 3.12). Pero en virtud de (3), $D_{n_1, n_2} = \sup\limits_u |G_X^*(u) - G_Y^*(u)|$, así que la distribución D_{n_1, n_2} no depende de F de ningún modo.

Se puede hallar la distribución exacta de la estadística D_{n_1, n_2}. Por ejemplo, cuando $n_1 = n_2 = n$,

$$\mathbf{P}(nD_{n,n} \geqslant k) = 2(C_{2n}^n)^{-1} \sum_{j=1}^{[n/k]} (-j)^{j+1} C_{2n}^{n-jk}, \tag{4}$$

$k = 1, 2, \ldots, n$. Este hecho fue establecido por Gnedendo y Koroliuk reduciendo esta tarea al simple problema de vagancias aleatorias (véase [32]).

En el § 1.6 hemos visto que la distribución $n_1 G_X^*(u)$ coincide con la distribución del proceso poissoniano $\zeta_1(u)$ a condición de que $\zeta_1(1) = n_1$. Como $G_X^*(u)$ y $G_Y^*(y)$ son independientes, la distribución $G_X^*(u) - G_Y^*(u)$, $u \in [0, 1]$ coincide con la distribución del proceso poissoniano compuesto $\zeta(u)$, en el que, con intensidad n_1, se producen saltos de magnitud $1/n_1$, y con intensidad n_2, saltos de magnitud $1/n_2$; la distribución ha de tomarse a condición de que ocurrieron $n_1 + n_2$ saltos y que $\zeta(1) = 0$. Por eso

$$\mathbf{P}(D_{n_1,\,n_2} < x) = \mathbf{P}(\sup_{u \leqslant 1} |\zeta(u)| < x / \zeta(1) = 0;\ \textit{ocurrieron } n_1 + n_2 \textit{ saltos}).$$

A base de este hecho, en el Suplemento II, además del teorema 1.6.2 de convergencia del proceso $w_n(u) = \sqrt{n_1}(G_X^*(u) - u)$ hacia el puente browniano $w^\circ(u)$, también se demuestra la afirmación de que hacia el referido puente también converge el proceso

$$w_{n_1,\,n_2}(u) = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}\,(G_X^*(u) - G_Y^*(u)).$$

Mejor dicho, para cualquier funcional f medible y continua en una métrica uniforme, la distribución $f(w_{n_1,\,n_2})$ converge hacia la distribución $f(w^\circ)$. De aquí se deduce inmediatamente la siguiente afirmación denominada teorema de Smirnov.

Teorema 1.

$$\lim_{n_1 \to \infty,\ n_2 \to \infty} \mathbf{P}\left(\sqrt{\frac{n_1 n_2}{n_1 + n_2}}\, D_{n_1,\,n_2} < x\right) = \mathbf{P}\left(\sup_{u \leqslant 1} |w^\circ(u)| < x\right) = K(x),$$

donde $K(x)$ es la función de Kolmogórov (véanse los §§ 1.8 y 3.12).

Como la función $K(x)$ está tabulada, el teorema 1 ofrece un medio cómodo para el cálculo aproximado del nivel de significación del criterio de Kolmogórov — Smirnov.

Le dejamos al lector que el mismo se cerciore de que el criterio de Kolmogórov — Smirnov es conciliable.

3. Criterio de signos. Sea $n_1 = n_2 = n$. Entonces, de las observaciones de las muestras X e Y se pueden componer n diferencias:

$$\mathrm{x}_1 - \mathrm{y}_1,\ \ldots,\ \mathrm{x}_n - \mathrm{y}_n. \tag{5}$$

Si es cierta la hipótesis H_1 y $\mathbf{P}_1 \times \mathbf{P}_1(\mathrm{x}_1 - \mathrm{y}_1 = 0) = 0$ para todas las $\mathbf{P}_1 \in \mathscr{P}$ (esto, evidentemente, siempre es así cuando $\mathscr{P}$ es un conjunto de distribuciones continuas), entonces

$$\mathbf{P}_1 \times \mathbf{P}_1(\mathrm{x}_1 - \mathrm{y}_1 > 0) = \mathbf{P}_1 \times \mathbf{P}_1(\mathrm{x}_1 - \mathrm{y}_1 < 0) = 1/2.$$

La estadística ν del criterio de signos es el número de diferencias positivas en (5)*⁾. El propio criterio se puede construir adoptando en calidad de conjunto crítico,

$$\Omega = \left\{(X,\ Y):\ \left|\nu - \frac{n}{2}\right| > c\right\}.$$

Como la distribución de ν no depende de $\mathbf{P}_1$,

$$\mathbf{P}_1 \times \mathbf{P}_1(\nu = k) = C_n^k 2^{-n},$$

por lo tanto, este criterio no es paramétrico.

El número c, según el nivel dado $1 - \varepsilon$ del criterio, se elige de la relación

$$k:|2k - n| \leqslant 2c\, C_n^k 2^{-n} \geqslant 1 - \varepsilon. \tag{6}$$

Como aquí el primer miembro crece de un modo discreto con el aumento de c, en calidad de solución conviene tomar el valor mínimo de c, con el que el primer miembro en (6) supera el valor de $1 - \varepsilon$.

Vemos que aquí se utiliza el criterio para verificar la hipótesis de que la probabilidad de éxito en el esquema de Bernoulli es igual a 1/2. Desde el punto de vista del problema inicial, se verifica no la hipótesis de homogeneidad, sino una hipótesis más amplia acerca de que

$$\mathbf{P}_1 \times \mathbf{P}_2(\mathrm{x}_1 - \mathrm{y}_1 < 0) = \int F_1(t)dF_2(t) = 1/2, \tag{7}$$

donde F_i corresponde a $\mathbf{P}_i$, $i = 1, 2$. La relación (7) significa que la mediana de distribución $\mathrm{x}_1 - \mathrm{y}_1$ es igual a 0.

El criterio de los signos del nivel asintótico $1 - \varepsilon$ tendrá la forma siguiente:

$$\pi(X,\ Y) = 1, \quad si \quad \frac{2\left|\nu - \frac{n}{2}\right|}{\sqrt{n}} > \lambda_{\varepsilon/2}, \tag{8}$$

$$\Phi_{0,1}(-\lambda_{\varepsilon/2},\ \lambda_{\varepsilon/2}) = 1 - \varepsilon.$$

Este criterio no es conciliable, ya que para $\mathbf{P}_1 \neq \mathbf{P}_2$ que satisfacen (7), $\beta_n(\mathbf{P}_1,\ \mathbf{P}_2) \to \varepsilon < 1$ cuando $n_1 \to \infty$, $n_2 \to \infty$.

4. Criterio de Wilkoxon. Este criterio se aplica ampliamente al verificar las hipótesis de homogeneidad.

Juntemos las muestras X e Y en una sola muestra $(X,\ Y)$ y construyamos de ella una serie variacional, o sea, situemos todas las observaciones

---

*⁾ Si en las muestras X e Y, debido al valor aproximado de los datos, resulta que algunas diferencias $\mathrm{x}_i - \mathrm{y}_i = 0$, entonces, éstas deben ser simplemente omitidas, tomando en calidad de n el número de diferencias distintas del cero.

en orden de crecimiento. Obtendremos una sucesión de tipo

$$\mathrm{y}^{(1)},\ \mathrm{y}^{(2)},\ \mathrm{x}^{(3)},\ \mathrm{y}^{(4)},\ \mathrm{x}^{(5)},\ \ldots, \tag{9}$$

donde el índice superior designa el número de observación en la serie variacional general, mientras que la letra indica la pertenencia a la muestra. Supongamos que $r_1, r_2, \ldots, r_{n_1}$ designan los números de elementos de la muestra X en la serie variacional (9). Para la sucesión escrita en (9), $r_1 = 3$, $r_2 = 5$. Llámase *estadística de Wilkoxon* la función

$$U = U(X,\ Y) = \sum_{i=1}^{n} (r_i - i),$$

donde $r_i - i$ es el número de elementos de la muestra Y que son menores de $\mathrm{x}_{(i)}$.

En vista de que el orden de observaciones en (9) es invariante respecto a las transformaciones monótonas de las variables (el orden de $F_X^*(t)$, $F_Y^*(t)$ será el mismo que para $F_X^*(F^{-1}(t))$, $F_Y^*(F^{-1}(t))$, donde F es la función de distribución), el criterio construido según la estadística U no será paramétrico.

Teorema 2. *Supongamos que $X \in \mathbf{P}_1$, $Y \in \mathbf{P}_2$ y $F_i \in \mathcal{F}$ son las funciones de distribución correspondientes a $\mathbf{P}_i$, $i = 1, 2$; $\mathcal{F}$ es la clase de todas las funciones de distribución continuas. Supongamos también, que $a = n_1/(n_1 + n_2) \to a_0$ cuando $n_1 \to \infty$, $n_2 \to \infty$. Entonces*

$$\frac{U - n_1 n_2 \mathbf{M} F_2(\mathrm{x}_1)}{\sqrt{n_1 n_2 (n_1 + n_2)}} \in \Phi_{0,\ \sigma^2}, \tag{10}$$

donde $\sigma^2 = (1 - a_0)\mathbf{D}F_2(\mathrm{x}_1) + a_0 \mathbf{D}F_1(\mathrm{y}_1)$.

Si $F_1 = F_2 = F$, *entonces* $F_2(\mathrm{x}_1) \in \mathbf{u}_{0,1}$, $F_1(\mathrm{y}_1) \in \mathbf{U}_{0,1}u$, *por consiguiente,* $\mathbf{M}F_2(\mathrm{x}_1) = 1/2$, $\mathbf{D}F_2(\mathrm{x}_1) = \mathbf{D}F_1(\mathrm{y}_1) = 1/12$.

Por lo tanto, el criterio de Wilkoxon de nivel asintótico $1 - \varepsilon$ tendrá la forma siguiente:

$$\left| U - \frac{n_1 n_2}{2} \right| > \frac{\lambda_{\varepsilon/2}\sqrt{n_1 n_2 (n_1 + n_2)}}{2\sqrt{3}}, \tag{11}$$

$$\Phi_{0,\ 1(-\lambda_{\varepsilon/2},\ \lambda_{\varepsilon/2})} = 1 - \varepsilon.$$

De (10) se deduce que este criterio tiene por objeto principal la verificación de la hipótesis (compárese con (7))

$$\int F_2(t) dF_1(t) = 1/2 \text{ o bien } \int (F_2(t) - F_1(t) dF_1(t) = 0. \tag{12}$$

Si admitimos, sin limitar la generalidad, que $F_1(t) \equiv t$, $t \in [0, 1]$, y si suponemos que $F_2(0) = 0$, $F_2(1) = 1$, entonces, en virtud de la igualdad

$$\int_0^1 (1 - F_2(t))dt = \mathbf{M}y_1,$$

la hipótesis que se verifica adoptará la forma $y_1 = 1/2$.

Esto significa que el criterio de Wilkoxon, al igual que el criterio de signos, es principalmente sensible a los desplazamientos de las distribuciones una respecto a otra. Para tales alternativas desplazadas, su potencia puede ser bastante grande (véase el ejemplo 1). Pero si $F_2 \neq F_1$ y se cumple (12), entonces, según el criterio de Wilkoxon, la hipótesis $\{F_2 = F_1\}$ será aceptada con una probabilidad próxima a $\Phi_{0,1}\left(-\frac{\lambda_{\varepsilon/2}}{2\sqrt{3}\sigma}, \frac{\lambda_{\varepsilon/2}}{2\sqrt{3}\sigma}\right)$. Esto significa que el criterio de Wilkoxon será inconciliable.

Demostración del teorema 2. La estadística U puede ser escrita de la forma siguiente:

$$U = \sum_{i=1}^{n_1} n_2 F_Y^*(x_i) = n_1 n_2 \int F_Y^*(t) dF_X^*(t).$$

Designemos

$$w_X(t) = \sqrt{n_1}(F_X^*(t) - F_1(t)), \quad w_Y(t) = \sqrt{n_2}(F_Y^*(t) - F_2(t)).$$

Entonces es evidente que

$$U = n_1 n_2 \int F_2(t) dF_1(t) + \sqrt{n_1 n_2 (n_1 + n_2)} \times$$
$$\times \left[\sqrt{a} \int w_Y(t) dF_1(t) + \sqrt{1 - a} \int F_2(t) dw_X(t)\right] + \sqrt{n_1 n_2} \int w_Y(t) dw_X(t). \tag{13}$$

Como aquí $\int F_2)t)dw_X(t) = \int w_X(t)dF_2(t)$ y, por consiguiente, las integrales segunda y tercera en (13) tienen la misma forma y son independientes, para demostrar el teorema es suficiente convencerse de que

$$\int w_Y(t) dF_1(t) \in \Phi_{0,\,\sigma_1^2}, \quad \sigma_1^2 = \mathbf{D}F_1(y_1) \tag{14}$$

y que

$$\frac{1}{\sqrt{n_1 + n_2}} \int w_Y(t) dw_X(t) \underset{P}{\to} 0. \tag{15}$$

En virtud del teorema 1.6.2,

$$\int w_Y(t) dF_1(t) \in \int w^\circ(F_2(t)) dF_1(t), \tag{16}$$

donde $w^\circ(u)$ es el puente borwniano. Para hallar la distribución de la última integral, señalaremos que las trayectorias del proceso wieneriano $w(u)$ de probabilidad 1 son continuas [11], $w^\circ(u) = w(u) - uw(1)$, y que, por lo tanto, la integral (16) es, por definición, el resultado de la convergencia casi segura de las sumas cuando $N \to \infty$,

$$\sum_{i=1}^{N} w(F_2(t_i))\Delta_i F_1 - m_1 w(1), \tag{17}$$

donde $m_1 = \int F_2(t)dF_1(t)$, $\{t_i\}_{i=0}^{N}$ forman la partición del eje real, $\Delta_i g = g(t_i) - g(t_{i-1})$, $w(F_2(t_i)) = \sum_{l=1}^{i} \Delta_l w(F_2)$, $w(1) = \sum_{l=1}^{N} \Delta_l w(F_2)$.

En virtud de la transformación de Abel,

$$\sum_{i=1}^{N}\left(\sum_{l=1}^{i} d_l\right)b_i = \sum_{l=1}^{N}\left(\sum_{i=l}^{N} b_i\right)d_l.$$

Por eso (17) es igual a

$$\sum_{l=1}^{N}(1 - F_1(t_{l-1}) - m_1)\Delta_l w(F_2). \tag{18}$$

Aquí $1 - m_1 = \int F_1(t)dF_2(t) = m_2$ y $\Delta_l w(F_2)$ son variables aleatorias normalmente distribuidas e independientes con parámetros $(0, \Delta_l F_2)$. Por eso la distribución (17), (18) será normal con media nula y con varianza

$$\sum_{l=1}^{N}(m_2 - F_1(t_{l-1}))^2\Delta_l F_2 \to \int (m_2 - F_1(t))^2 dF_2(t) = \mathbf{D}F_1(y_1).$$

La relación (14) queda demostrada.

Para demostrar (15)*), lo más fácil es estimar la varianza de la integral en (15). Volviendo a aproximar la integral con ayuda de la suma final, es posible convencerse que la varianza

$$D_{X,Y} = \mathbf{M}\left(\int w_Y(t)dw_X(t)\right)^2$$

está limitada cuando $n_1 \to \infty$, $n_2 \to \infty$. De aquí y de la desigualdad de Chébishev resulta (15). Debido a los cálculos voluminosos y rutinarios, omitiremos la demostración del carácter limitado de $D_{X,Y}$. ◁

Datos más exactos acerca de los criterios de signos y de Wilkoxon se exponen en [41].

Ejemplo 1. Hemos señalado que los criterios de signos y de Wilkoxon son los más sensibles a los desplazamientos. Por eso es interesante comparar su potencia con la del criterio óptimo en el problema donde la homogeneidad se verifica para la familia $\mathscr{P}$ de distribuciones que sólo se distinguen por sus desplazamientos. Pues, supongamos que

$$\mathscr{P} = \{\Phi_{\alpha,1}\}, \quad \mathbf{P}_1 = \Phi_{\alpha_1,1}, \quad \mathbf{P}_2 = \Phi_{\alpha_2,1}, \quad n_1 = n_2 = n.$$

En este caso, conforme al teorema 1.1, para verificar la hipótesis $H_1 = \{\mathbf{P}_1 = \mathbf{P}_2\} = \{\alpha_1 = \alpha_2\}$ frente a $H_2^b = \{|\alpha_1 - \alpha_2| \geqslant b/\sqrt{n}$ existe el criterio asintóticamente minimax π_0 de nivel $1 - \varepsilon$, que tiene la forma

$$|\bar{x} - \bar{y}| > \lambda_{\varepsilon/2}\sqrt{2/n}, \quad \Phi_{0,1}((-\lambda_{\varepsilon/2}, \lambda_{\varepsilon/2})) = 1 - \varepsilon$$

(el hecho de que en nuestro ejemplo esta desigualdad equivale a (1.3 y 1.4), el lector puede comprobarlo personalmente). Tomemos este criterio por

*) La integral en (15) converge respecto a la distribución hacia $\int w^0(F_2(t))dw^0(F_1(t))$.

patrón para la comparación con otros criterios y examinemos la alternativa $(\mathbf{P}_1, \mathbf{P}_2)$, donde $\alpha_2 = \alpha_1 + c/\sqrt{n}$ (examinamos las alternativas semejantes para no tratar el problema de grandes desviaciones). Es evidente que en este caso $(\overline{x} - \overline{y}) \in \Phi_{-c/\sqrt{n},\, 2/n}$. Por lo tanto,

$$\begin{aligned}\beta_{\pi_0}(\mathbf{P}_1, \mathbf{P}_2) = \mathbf{P}_1 \times \mathbf{P}_2(|\overline{x} - \overline{y}| > \lambda_{\varepsilon/2}\sqrt{2n}) =& \\ = 1 - \Phi_{-c/\sqrt{2},1}(-\lambda_{\varepsilon/2}, \lambda_{\varepsilon/2}) =& \\ = 1 - \Phi_{0,1}(-\lambda_{\varepsilon/2} + c/\sqrt{2}, \lambda_{\varepsilon/2} + c/\sqrt{2}) \equiv \beta_0(c).& \qquad (19)\end{aligned}$$

Examinemos ahora el criterio de signos (8), designándolo por π_1. Haciendo uso del desarrollo en serie de las potencias de $c/\sqrt{n}$, hallamos $(\Phi_{\alpha,\sigma^2}(x) = \Phi_{\alpha,\sigma^2}((-\infty, x)))$

$$\mathbf{P}_1 \times \mathbf{P}_2(x_1 - y_1 < 0) = \Phi_{0,2}\left(\frac{c}{\sqrt{n}}\right) = \frac{1}{2} + \frac{c}{\sqrt{n}} \cdot \frac{1}{2\sqrt{\pi}} + O\left(\frac{1}{n}\right).$$

Por eso en el punto $(\mathbf{P}_1, \mathbf{P}_2)$

$$\frac{2}{\sqrt{n}}\left(\nu - \frac{n}{2} + \frac{c\sqrt{n}}{2\sqrt{n}}\right) \in \Phi_{0,1}.$$

Por consiguiente, para el criterio de signos π_1 de nivel asintótico $1 - \varepsilon$,

$$\begin{aligned}\beta\pi_1(\mathbf{P}_1, \mathbf{P}_2) = \mathbf{P}_1 \times \mathbf{P}_2\left(2\left|\nu - \frac{n}{2}\right| > \lambda_{\varepsilon/2}\sqrt{n}\right) \to& \\ \to 1 - \Phi_{0,1}\left(\left(-\lambda_{\varepsilon/2} + \frac{c}{\sqrt{\pi}}, \lambda_{\varepsilon/2} + \frac{c}{\sqrt{\pi}}\right)\right).&\end{aligned}$$

Volvamos, por último, al criterio π_2 de Wilkoxon (véase (11)) que en nuestro caso tiene la forma

$$\left|U - \frac{n^2}{2}\right| > \frac{\lambda_{\varepsilon/2} n^{3/2}}{\sqrt{6}}.$$

Evidentemente, la estadística U es invariante respecto a la transformación de desplazamiento de los elementos de las muestras X e Y. Por eso se puede considerar que $\mathbf{P}_1 = \Phi_{0,1}$, $\mathbf{P}_2 = \Phi_{c/\sqrt{n},1}$ y, por lo tanto,

$$\begin{aligned}\mathbf{M}F_2(x_1) = \int F_2(t)dF_1(t) = \int \Phi_{0,1}\left(t - \frac{c}{\sqrt{n}}\right) d\Phi_{0,1}(t) =& \\ = \Phi_{0,2}\left(-\frac{c}{\sqrt{n}}\right) = \frac{1}{2} - \frac{c}{\sqrt{n}} \cdot \frac{1}{2\sqrt{\pi}} + O\left(\frac{1}{n}\right).&\end{aligned}$$

Como $\mathbf{D}F_2(x_1) \to \mathbf{D}F_1(x_1) = 1/12$, $\mathbf{D}F_1(y_1) \to \mathbf{D}F_1(x_1) = 1/12$, según el teorema 2,

$$\beta_{\pi_2}(\mathbf{P}_1, \mathbf{P}_2) = \mathbf{P}_1 \times \mathbf{P}_2\left(\left|U - \frac{n^2}{2}\right| > \frac{\lambda_{\varepsilon/2} n^{3/2}}{\sqrt{6}}\right) =$$
$$= 1 - \mathbf{P}_1 \times \mathbf{P}_2\Bigg(-\lambda_{\varepsilon/2} + c\sqrt{\frac{3}{2\pi}} \leqslant$$
$$\leqslant \sqrt{6}n^{-3/2}\left(U - \frac{n^2}{2} + \frac{n^{3/2}c}{2\sqrt{\pi}}\right) \leqslant \lambda_{\varepsilon+2} + c\sqrt{\frac{3}{2\pi}}\Bigg) =$$
$$= 1 - \Phi_{0,1}\left(-\lambda_{\varepsilon/2} + c\sqrt{\frac{3}{2\pi}}, \lambda_{\varepsilon/2} + c\sqrt{\frac{3}{2\pi}}\right).$$

Ahora debemos señalar que $\beta_0(c)$ (véase (19)) es una función monótona creciente de c y que, con grandes valores de n,

$$\beta\pi_1(\mathbf{P}_1, \mathbf{P}_2) \approx \beta_0\left(\sqrt{\frac{2}{\pi}}\, c\right), \ \beta\pi_2(\mathbf{P}_1, \mathbf{P}_2) \approx \beta_0\left(\sqrt{\frac{3}{\pi}}\, c\right).$$

Ahora bien, para cada $c > 0$, el más potente entre los π_0, π_1 y π_2 resulta, como era de esperar, el criterio π_0. Le siguen el criterio de Wilkoxon y el de signos; con la particularidad de que el criterio de Wilkoxon cede muy poco al criterio π_0, ya que $\sqrt{3/\pi} \approx 0{,}977$.

Si para ese mismo desplazamiento $\alpha_2 - \alpha_1 = c/\sqrt{n}$ examinamos las muestras X' e Y' de nivel $n' > n$, entonces, para obtener (con ayuda de los cálculos efectuados) la potencia de los criterios $\pi_j(X', Y')$ en el punto $(\mathbf{P}_1, \mathbf{P}_2)$, debemos examinar el problema anterior para un nuevo valor de c, igual a $c' = c\sqrt{n'}/\sqrt{n}$ (entonces $\alpha_2 - \alpha_1$ puede escribirse en forma de $c'/\sqrt{n'}$). Por consiguiente, las potencias de $\pi_1(X', Y')$ y de $\pi_2(X', Y')$ en ese mismo punto $(\mathbf{P}_1, \mathbf{P}_2)$ serán aproximadamente iguales a

$$\beta_0\left(\sqrt{\frac{2}{\pi}}\, c'\right) = \beta_0\left(\sqrt{\frac{2n'}{\pi n}}\, c\right), \ \beta_0\left(\sqrt{\frac{3}{\pi}}\, c'\right) = \beta_0\left(\sqrt{\frac{3n'}{\pi n}}\, c\right).$$

Igualando $\frac{2n'}{\pi n} = 1$, $\frac{3n'}{\pi n} = 1$, obtenemos los valores de $n' = \frac{\pi}{2}n$, $n' = \frac{\pi}{3}n$ (estos valores no dependen de c) para el número de observaciones que necesitamos realizar a fin de obtener con ayuda de los criterios π_1 y π_2, respectivamente, la misma potencia que para el criterio π_0 con n observaciones. Por ejemplo, para $n = 100$ observaciones con criterio π_0 necesitaremos, para obtener esos mismos resultados, $n' \approx 105$ observaciones con criterio π_2 y $n' \approx 157$ observaciones con criterio π_1.

Obtendríamos absolutamente otros resultados si hubiéramos verificado la homogeneidad para la familia $\mathscr{P} = \{\Phi_{0,\sigma^2}\}$. En este caso los criterios de signos y de Wilkoxon resultarían inconciliables. Más aún, el criterio de signos de nivel $1 - \varepsilon$ sería, en realidad, equivalente al criterio $\pi \equiv \varepsilon$ que no depende de las muestras, ya que $\mathbf{M}(x_1 - y_1) = 0$ y $\mathbf{P}_1 \times \mathbf{P}_2(x_1 - y_1 > > 0) = 1/2$ para cualquier par de distribuciones $\mathbf{P}_1$ y $\mathbf{P}_2$ de $\mathscr{P}$. Para este problema se podrían examinar otros criterios no paramétricos que utilizan las estadísticas r_i, por ejemplo, el criterio $\sum_{i=0}^{n_1} (r_{i+1} - r_i)^2$, $r_0 = 0$, $r_{n_1+1} = = n_2$ que se asemeja por sus propiedades al criterio de Morán (§ 3.12).

5. Criterio χ^2 como criterio asintóticamente óptimo para verificar la homogeneidad según los datos agrupados. En este apartado supondremos que los datos en ambas muestras X e Y de volúmenes n_1 y n_2, respectivamente, están agrupados (véase el § 3.16). En este caso en vez de las muestras X e Y es posible utilizar los vectores $\nu = (\nu_1, \ldots, \nu_r)$ y $\mu = (\mu_1, \ldots, \mu_r)$ de las frecuencias de observaciones de las muestras X e Y, respectivamente, que cayeron en los intervalos $\Delta_1, \ldots, \Delta_r$ que definen la agrupación. Designemos por $\theta_i = (\theta_{i1}, \ldots, \theta_{ir})$, $i = 1, 2$, los vectores de las probabilidades de que las observaciones de la primera y la segunda muestras caigan en los intervalos $\Delta_1, \ldots, \Delta_r$, de modo que $\theta_{1l} = \mathbf{P}(x_j \in \Delta_l)$, $\theta_{2l} = \mathbf{P}(y_j \in \Delta_l)$. Las muestras aproximadas X e Y entonces pueden considerarse como muestras de las familias paramétricas $\mathbf{B}_{\theta_1}$ y $\mathbf{B}_{\theta_2}$, respectivamente. Ahora bien, el problema llega a ser paramétrico y podemos utilizar los resultados citados en el ejemplo 1 del párrafo precedente. De este ejemplo se deduce que si verificamos la hipótesis de homogeneidad $H_1 = \{\theta_1 = \theta_2\}$ en el caso en que el parámetro θ está localizado, o sea, los valores de θ_1 y θ_2 se sitúan en el entorno del punto $\theta_0 = (\theta_{01}, \ldots, \theta_{0r})$, entonces el criterio asintóticamente minimax de nivel asintótico $1 - \varepsilon$ para verificar H_1 frente a

$$H_2^b = \left\{ \sum_{l=1}^{r} \frac{(\theta_{1l} - \theta_{2l})^2}{\theta_{0l}} \geqslant \frac{b^2}{n_2} \right\}$$

tiene la forma

$$\sum_{l=1}^{r} \left(\frac{\nu_l}{n_1} - \frac{\mu_l}{n_2} \right)^2 \frac{n_1 n_2}{\nu_l + \mu_l} \geqslant h_\varepsilon,$$

donde h_ε es una cuantila del orden de $1 - \varepsilon$ de la distribución χ^2 con $r - 1$ grados de libertad. Este es precisamente el criterio χ^2 para verificar la homogeneidad según los datos agrupados.

En calidad de criterio asintóticamente equivalente puede ser considerado el criterio

$$\sum_{l=1}^{r} \nu_l \ln \frac{\nu_l}{n_1} + \sum_{l=1}^{r} \mu_l \ln \frac{\mu_l}{n_2} - \sum_{l=1}^{r} (\nu_l + \mu_l) \ln \frac{\nu_l + \mu_l}{n_1 + n_2} > \frac{h_\varepsilon}{2}.$$

§ 3. Problemas de regresión

1. Planteamiento del problema. En las aplicaciones a menudo surgen problemas referentes a las observaciones cuya distribución varía en distintos experimentos al cambiar algunos parámetros que caracterizan estos últimos. El conjunto de valores de los parámetros mencionados en el i-ésimo experimento, $i = 1, \ldots, n$ lo designaremos por

$$\mathrm{x}_i = (\mathrm{x}_{i,1}, \ldots, \mathrm{x}_{i,r})$$

(así que r es la dimensión de los vectores x_i). Los valores de $\mathrm{x}_{i,k}$ son determinados por el experimentador o por la naturaleza del fenómeno que se estudia. Designemos el vector $(\mathrm{x}_{1,k}, \ldots, \mathrm{x}_{n,k})$ por la letra X_k, y la matriz $\left(\frac{X_1}{X_r}\right) = (x_1^T, \ldots, x_n^T)$, por la letra X. Ahora bien, aquí, a distinción de lo expuesto anteriormente, X es una matriz del orden de $r \times n$ y puede ser un conjunto no aleatorio arbitrario de números cuya naturaleza no nos interesará. El vector de observaciones se designa por $Y = (\mathrm{y}_1, \ldots, \mathrm{y}_n)$.

Los problemas de regresión están relacionados con la suposición de que las observaciones y_i, en función del conjunto de parámetros $\mathrm{x}_i = (\mathrm{x}_{i,1}, \ldots, \mathrm{x}_{i,r})$, tienen la forma

$$\mathrm{y}_i = \alpha_1 \mathrm{x}_{i,1} + \ldots + \alpha_r \mathrm{x}_{i,r} + \xi_i, \quad i = 1, \ldots, n, \tag{1}$$

donde $\alpha = (\alpha_1, \ldots, \alpha_r)$ son constantes desconocidas para nosotros, y $\xi_i \in \Phi_{0,\sigma^2}$ son constantes independientes.

La constante α_1 desempeña a menudo un papel especial, ya que en una serie de casos ésta separa en la representación (1) el sumando constante, lo cual corresponde a que en la matriz X se supone de antemano $X_1 = (1, \ldots, 1)$ $(\mathrm{x}_{i,1} \equiv 1)$. No haremos uso de esta suposición. Las variables aleatorias ξ_i se deben a los ruidos y fluctuaciones o a los errores de medición.

En forma matricial las relaciones (1) pueden escribirse del modo siguiente

$$Y = \alpha X + \xi. \tag{2}$$

La regresión que tiene la forma (1) y (2) se llama lineal (tanto respecto a α como respecto a X). En calidad de problemas de regresión pueden considerarse tanto el problema de estimación de los parámetros desconocidos α y σ^2, si se sabe que es válida (1), (2), como el problema de verificación de la propia hipótesis de que la representación (1), (2) tiene lugar. En ambos casos, como datos iniciales sirve la «muestra» (X, Y). El término «muestra» se utiliza aquí en un sentido más amplio que antes, designando con él el conjunto de resultados de observaciones que no tienen obligatoriamente la misma naturaleza. Además, recordemos que la primera de las dos «muestras» X e Y puede ser no aleatoria. La matriz X se llama, a veces, *regresor* y el vector Y, *respuesta*.

El modelo de regresión (1), (2) es muy general si se tiene en cuenta que y_i depende del conjunto de parámetros. Suponiendo, por ejemplo, $x_{i,k} = \psi_k(z_i)$, donde $\psi_1, \ldots, \psi_r$ es un conjunto dado de funciones, y z_i son los valores del parámetro unidimensional, obtenemos el modelo

$$y_i = \alpha_1\psi_1(z_i) + \ldots + \alpha_r\psi_r(z_i) + \xi_i, \quad i = 1, \ldots, n, \tag{3}$$

de la regresión respecto a las funciones arbitrarias $\psi_1, \ldots, \psi_r$ (y, como antes, lineal respecto a α). Si $\psi_1(z) \equiv 1$, $\psi_2(z) \equiv z$ y $r = 2$, obtenemos el modelo de una *regresión lineal* elemental (unidimensional) (fig. 6).

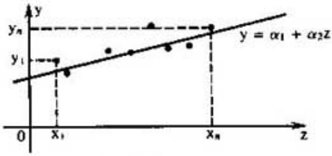

Fig. 6.

A distinción del modelo elemental, el modelo general (1), (2) se denomina, a veces, regresión *múltiple*. En general, como vemos, los problemas de regresión están relacionados con el estudio (existencia) de la dependencia funcional $y = \varphi(x)$ para una clase dada de funciones φ en los casos en que las observaciones de la variable y, para x dada, van acompañadas de «ruidos» en forma de desviaciones aleatorias.

Las filas $X_1, \ldots, X_r$ de la matriz X en (2) suelen elegirse de modo que sean linealmente independientes (de otro modo no podremos estimar las coordenadas de α). También seguiremos este convenio que significa que el rango de la matriz X es igual a r.

A veces es más cómodo tratar con los vectores ortogonales $X_1, \ldots, X_r$, o sea, con los vectores que satisfacen la condición $(X_i, X_j) = 0$, $i \neq j$, donde (a, b) significa el producto escalar. Si el conjunto inicial de vectores linealmente independientes (X_k) no posee tal propiedad, el mismo puede ser ortogonalizado introduciendo nuevos vectores:

$$\begin{aligned} X_1' &= X_1, \\ X_2' &= X_2 + a_{2,1}X_1, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ X_r' &= X_r + a_{r,r-1}X_{r-1} + \ldots + a_{r,1}X_1. \end{aligned} \tag{4}$$

Los coeficientes $a_{k,j}$ se deducen fácilmente de las condiciones de ortogonalidad $X_k' \perp X_j'$, $k \neq j$, así que, por ejemplo, $a_{2,1} = -\dfrac{(X_1, X_2)}{(X_1, X_1)}$. Las relaciones (4) pueden ser escritas en forma de $X' = AX$, donde A es una matriz invertible triangular (con unidades que pasan por la diagonal principal). De aquí obtenemos $X = A^{-1}X'$, $Y = \alpha A^{-1}X' + \xi$. Hemos llegado al problema de regresión con coeficientes $\beta = \alpha A^{-1}$. El vector α se reconstruye de un modo evidente por β con ayuda de la igualdad $\alpha = \beta A$.

Para una regresión lineal elemental, la suposición acerca de la ortogonalidad de $X_1 = (1, \ldots, 1)$ y $X_2 = (z_1, \ldots, z_n)$ significa la suposición de $\sum z_i = 0$ que, evidentemente, puede ser satisfecha variando el comienzo de la lectura de la variable z.

2. Estimación de los parámetros. En lo sucesivo supondremos por doquier, que $r < n$ y que los vectores X_k, $k = 1, \ldots, r$, son linealmente independientes. La función de verosimilitud de la observación Y (con X dada) para la regresión (1), (2) es igual a

$$f_{\alpha,\sigma^2}(Y) = \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n_{i=1} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \sum_{k=1}^{r}\alpha_k x_{i,k}\right)\right\}^2 =$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n \exp\left\{-\frac{|Y - \alpha X^2|}{2\sigma^2}\right\}. \qquad (5)$$

La función (5) depende del parámetro $\theta = (\alpha, \sigma^2)$. Nótese que si (5) se considera como función de verosimilitud no de una sola observación Y (o, (X, Y), sino de n observaciones $y_1, \ldots, y_n$, ella no corresponderá a la muestra de una familia paramétrica cualquiera. Las observaciones y_i se refieren a distintas distribuciones Φ_{γ_i,σ^2}, $\gamma_i = \sum_{k=1}^{r} \alpha_k x_{i,k}$ que dependen de x_i. Por eso las consideraciones expuestas en los capítulos anteriores, donde se utilizó la misma distribución de los elementos de la muestra, aquí no se aplican directamente.

Así pues, examinaremos (5) como función de verosimilitud de la observación (X, Y). Hagamos uso del método de verosimilitud máxima. Directamente de (5) se deduce que la estimación de verosimilitud máxima $\alpha^* = \hat{\alpha}^*$ que maximiza $f_\theta(Y)$ respecto a α es la estimación que minimiza $|Y - \alpha X|^2$. Por eso en nuestro caso el método de verosimilitud máxima coincide con el «método de cuadrados mínimos».

Designemos por $\mathscr{L}[X]$ el subespacio tendido en los vectores $X_1, \ldots, X_r$. El mismo constituye una población de puntos en forma de αX cuando α recorre los valores de R^r. La dimensión de este espacio es r y en él sólo hay un punto $\beta = \alpha^* X$ que es el menos alejado de Y (fig. 7). El valor de β está unívocamente determinado por la condición de ortogonalidad $Y - \beta$ y $\mathscr{L}[X]$, o bien, que es lo mismo, por las r condiciones

$$(Y - \alpha^* X, X_k) = (Y - \alpha^* X)X_k^T = 0, \ k = 1, \ldots, r.$$

En forma matricial estas condiciones pueden escribirse del modo siguiente: $(Y - \alpha^* X)X^T = 0$. De aquí hallamos

$$\alpha^* = YX^T(XX^T)^{-1}. \tag{6}$$

Aquí, la matriz inversa $(XX^T)^{-1}$ (del orden de $r \times r$) existe, ya que la matriz $D = XX^T$ está definida positivamente. En efecto, hemos visto que existe

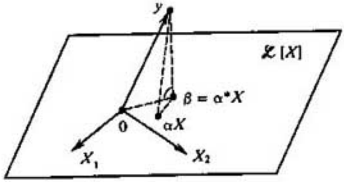

Fig. 7.

una matriz no degenerada A tal que las filas de la matriz $X' = AX$ son ortogonales. Por consiguiente, la matriz D puede ser escrita del modo siguiente:

$$XX^T = A^{-1}X'(X')^T(A^{-1})^T = A^{-1}B(A^{-1})^T,$$

donde $B = X'(X')^T$ es una matriz diagonal con los elementos

$$(X_i', X_j') = \begin{cases} |X_i'|^2 > 0 & \text{para } i = j, \\ 0 & \text{para } i \neq j. \end{cases}$$

Por lo tanto, B está definida positivamente, $aBa^T > 0$ para cualquier $a \in R^r$, $a \neq 0$. Poniendo $b = aA$, obtenemos $bDb^T = aAXX^TA^Ta^T = aBa^T > 0$ para cualquier $b \in R^r$, $b \neq 0$, que es lo que se necesitaba demostrar.

Si X_k son ortogonales, de (6) hallamos $\alpha_k^* = \dfrac{(Y, X_k)}{(X_k, X_k)}$.

El resultado (6) también puede ser obtenido derivando (5) respecto a α_k e igualando a cero las derivadas.

La diferencia $Y - \alpha^* X$ a veces se llama *resto*. Esta diferencia es ortogonal a $\mathscr{L}[X]$ y, al mismo tiempo, a cualquier vector $\gamma X \in \mathscr{L}[X]$, $\gamma \in R^r$. Si se adopta $\gamma = \alpha^* - \alpha$, de la igualdad $Y - \alpha X = Y - \alpha^* X + (\alpha^* - \alpha)X$ se deducirá

$$|Y - \alpha X|^2 = |Y - \alpha X|^2 + |(\alpha^* - \alpha)X|^2. \tag{7}$$

Hallemos ahora la e.v.m. para σ^2. De (5) se deduce que ésta será la misma estimación que para una familia normal (se puede volver a derivar (5) respecto a σ, igualando a cero la derivada), así que

$$(\widehat{\sigma^2})^* = \frac{1}{n}|Y - \alpha^* X|^2. \tag{8}$$

Pongamos

$$(\sigma^2)^* = \frac{1}{n-r}\,|Y - \alpha^* X|^2 = \frac{1}{n-r}\,(\hat{\sigma}^2)^*. \tag{9}$$

En lo sucesivo E_l significará una matriz unidad de orden l, $\sigma^* = \sqrt{(\sigma^2)^*}$.

Teorema 1. (6) y (9) *son las estimaciones eficientes no desplazadas e independientes de los parámetros* α y σ^2. *Además,*

$$(\alpha^* - \alpha)D^{1/2} \in \Phi_{0,\sigma^2 E_r},\ D = XX^T, \tag{10}$$

$$(n-r)(\sigma^2)^*/\sigma^2 = |Y - \alpha^* X|^2/\sigma^2 \in H_{n-r}. \tag{11}$$

Si X_k son ortogonales, α_k^* son independientes,

$$(\alpha_k^* - \alpha_k)|X_k| \in \Phi_{0,\sigma^2}. \tag{12}$$

Corolario 1. *De* (10) y (11) *se deduce que*

$$\frac{(\alpha^* - \alpha)D(\alpha^* - \alpha)^T}{(n-r)(\sigma^2)^*} = \frac{|(\alpha^* - \alpha)X|^2}{|Y - \alpha^* X|^2} \in F_{r,n-r}. \tag{13}$$

Sean $\overline{\alpha}$, $\overline{\alpha}^*$ "subvectores" de dimensión $l \leqslant r$ de los vectores α y α^* formados por coordenadas de números fijos $k_1, \ldots, k_l$, y sea $\overline{X}$ una matriz formada por las filas $X_{k_1}, \ldots, X_{k_l}$. Entonces, si X_k, $k = 1, \ldots, r$, son ortogonales, entonces

$$(\overline{\alpha}^* - \overline{\alpha})(\overline{X}X^T)^{1/2} \in \Phi_{0,\sigma^2 E_l},\ (\alpha_k^* - \alpha_k)|X_k|/\sigma^* \in \mathbf{T}_{n-r}. \tag{14}$$

Demostración del teorema 1. En vista de que $YX^T = \alpha XX^T + \xi X^T$, entonces

$$\alpha = (YX^T - \xi X^T)D^{-1},\quad \alpha^* - \alpha = \xi X^T D^{-1}. \tag{15}$$

La matriz de segundos momentos del vector $(\alpha^* - \alpha)D^{1/2}$ es igual a

$$\mathbf{M}D^{1/2}(\alpha^* - \alpha)^T(\alpha^* - \alpha)D^{1/2} =$$
$$= D^{1/2}D^{-1}X\mathbf{M}\xi^T\xi X^T D^{-1}D^{1/2} = \sigma^2 E_r.$$

Como las componentes de este vector son normales, ellas son independientes y $\dfrac{1}{\sigma^2}\,|(\alpha^* - \alpha)D^{1/2}|^2 \in \mathbf{H}_r$. Luego, en virtud de (7) y (9),

$$(n-r)(\sigma^2)^* = |Y - \alpha^* X|^2 = |\xi|^2 - |(\alpha^* - \alpha)X|^2.$$

Cerciorémonos ahora de que los vectores α^* e $Y - \alpha^* X$ (y, por consiguiente, α^* y σ^*) son independientes. En virtud de su normalidad es suficiente comprobar que los coeficientes de correlación entre sus componentes son iguales a cero o bien, que es lo mismo, que la matriz de segundos momentos centrales $\mathbf{M}(\alpha^* - \alpha)^T(Y - \alpha^* X)$ es igual a cero. Nótese que en virtud de (6),

$$\alpha^* X = YX^T(XX^T)^{-1}X = YX^T D^{-1}X,$$

y el vector $\alpha^* X$ se obtiene de Y mediante la proyección de Y sobre $\mathscr{L}[X]$. El operador de proyección, definido por la matriz $\Pi = X^T D^{-1} X$, posee propiedades evidentes: $\Pi^2 = \Pi$, $BX\Pi = BX$ para cualquier matriz B que tiene r filas. Por eso, en virtud de (15),

$$\mathbf{M}(\alpha^* - \alpha)^T (Y - \alpha^* X) = \mathbf{M} D^{-1} X \xi^T (\xi - \xi X^T D^{-1} X) =$$
$$= D^{-1} X \sigma^2 (E_n - \Pi) = 0.$$

Demostremos ahora (11). En virtud de (7),

$$|Y - \alpha^* X|^2 = |\xi|^2 - |(\alpha^* - \alpha) X|^2 = |\xi|^2 - |(\alpha^* - \alpha) D^{1/2}|^2,$$

donde $\dfrac{1}{\sigma^2} |\xi|^2 \in \mathbf{H}_n$, $\dfrac{1}{\sigma^2} |(\alpha^* - \alpha) D^{1/2}|^2 \in \mathbf{H}_r$ (véase (10)). La afirmación (11) será el corolario de estas relaciones y del lema 1.

Lema 1. *Si $\eta = \eta_1 + \eta_2$, donde η_1 y η_2 son independientes, $\eta \in \mathbf{H}_n$, $\eta_1 \in \mathbf{H}_r$, entonces $\eta_2 \in \mathbf{H}_{n-r}$.*

Demostración. Si se designa por $\varphi(t)$ la función característica de la distribución $\mathbf{H}_1$: $\varphi(t) = (1 + 2it)^{-1/2}$, entonces

$$\mathbf{M} e^{i\eta t} = \varphi(t)^n = \varphi(t)^r \mathbf{M} e^{i\eta_2 t}.$$

Como $\varphi(t) \neq 0$ en el eje real, entonces $\mathbf{M} e^{i\eta_2 t} = \varphi(t)^{n-r}$. El lema queda demostrado.

El no desplazamiento de las estimaciones α^* y $(\sigma^2)^*$ se deduce con evidencia de (10), (11) ($\mathbf{M}\eta = l$ si $\eta \in \mathbf{H}_l$).

Nos queda demostrar la eficacia de la estimación $\theta^* = (\alpha^*, (\sigma^2)^*)$. Para esto debemos notar que la familia (5) pertenece al tipo exponencial, ya que (5) es representable en la forma (véase (2.15.1))

$$f_\theta(Y) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\left(|Y|^2 - 2(Y, \alpha X) + |\alpha X|^2\right)\right\} =$$
$$= h(Y) \exp\left\{\sum_{k=1}^{r+1} a_k(\theta) U_k(Y) + V(\theta)\right\},$$

donde

$$h(Y) = (2\pi)^{-n/2}, \quad V(\theta) = -n \ln \sigma - \frac{1}{2\sigma^2} |\alpha X|^2,$$

$$a_k(\theta) = \frac{\alpha_k}{\sigma^2}, \quad U_k(Y) = (Y, X_k), \quad k = 1, \ldots, r,$$

$$a_{r+1}(\theta) = -\frac{1}{2\sigma^2}, \quad U_{r+1}(Y) = |Y|^2.$$

Como las condiciones de los teoremas 2.15.1 y 2.15.2 aquí se cumplen, la estadística $U = (U_1(X), \ldots, U_{r+1}(X))$ (y junto con ella también θ^*) es una estadística mínima suficiente completa. De aquí se desprende (véase el corolario 2.15.1) la eficacia de θ^*.

La afirmación (12) resulta con evidencia de (10), ya que para X_k ortogonales, la matriz $D^{1/2}$ es diagonal a los elementos $|X_k|$ dispuestos diagonalmente. El teorema queda demostrado.

Observación 1. Hotelling (véase [83] demostró que $\mathbf{D}\alpha_k \geqslant \sigma^2/|X_k|^2$ y la igualdad se alcanza tan sólo en el caso cuando X_k son ortogonales. Ahora bien, al planificar un experimento para valores dados de $|X_k|$, la elección óptima del regresor X consiste en hacer ortogonales X_k.

Observación 2. Es interesante comparar la matriz de segundos momentos de la estimación θ^*, con la frontera inferior para las estimaciones no desplazadas, la cual se define, en virtud de la desigualdad multidimensional de Rao—Cramer, por la matriz $I^{-1}(\theta)$, donde $I(\theta)$ es la matriz de información de Fisher:

$$I(\theta) = \|I_{ij}(\theta)\|, \; I_{ij}(\theta) = \mathbf{M}\theta \, \frac{\partial L}{\partial \theta_i} \cdot \frac{\partial L}{\partial \theta_j}, \; L = L(Y;\, \theta) = \ln f_\theta(Y).$$

Aquí hemos adoptado $\theta_k = \alpha_k$, $k = 1, \ldots, r$, $\theta_{r+1} = \sigma^2$. Supongamos, para abreviar, que X_k son ortogonales. De la independencia de θ_k^* se deduce que la matriz $\mathbf{M}_\theta(\theta^* - \theta)^T(\theta^* - \theta)$ será diagonal a los elementos dispuestos diagonalmente:

$$\mathbf{M}_\theta(\alpha_k^* - \alpha)^2 = \frac{\sigma^2}{|X_k|^2}, \quad k = 1, \ldots, r,$$

$$\mathbf{M}_\theta((\sigma^2) - \sigma^2) = \mathbf{M}\left(\frac{\sigma^2 \chi^2_{n-r}}{n - r} - \sigma^2\right)^2 = \frac{\sigma^4}{n - r}\, \mathbf{M}(\chi_1^2 - 1)^2 = \frac{2\sigma^4}{n - r},$$

donde $\chi_l^2 \in \mathbf{H}_l$.

Por otro lado, para la matriz $I(\theta)$, en virtud de que

$$\frac{\partial L}{\partial \alpha_k} = \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{r} \alpha_j x_{ij}\right) x_{ik} = \frac{1}{\sigma^2}\,(Y - \alpha X) X_k^T,$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{r} \alpha_j x_{ij}\right)^2 = \frac{1}{2\sigma^2}\left(\frac{|Y - \alpha X|^2}{\sigma^2} - n\right),$$

hallamos, cuando $k = 1, \ldots, r$:

$$I_{kk}(\theta) = \mathbf{M}_\theta \, \frac{1}{\sigma^4} \, X_k(Y - \alpha X)^T(Y - \alpha X)X_k^T =$$

$$= \frac{1}{\sigma^4}\, \mathbf{M} X_k \xi^T \xi X_k^T = \frac{1}{\sigma^4}\, \mathbf{M}(\xi,\, X_k)^2 = \frac{|X_k|^2 \mathbf{M}|\xi|^2}{\sigma^4} = \frac{|X_k|^2}{\sigma^2},$$

$$I_{r+1,r+1}(\theta) = \frac{1}{4\sigma^4}\, \mathbf{M}\left[\sum_{l=1}^{n} \left(\frac{\xi_l^2}{\sigma^2} - 1\right)\right]^2 = \frac{n}{2\sigma^4}, \; I_{ij}(\theta) = 0$$

cuando $i \neq j$. Así que

$$I^{-1}(\theta) = \begin{Bmatrix} \frac{\sigma^2}{|X_1|^2} & 0 & 0 \\ & \ddots & \vdots \\ 0 & \frac{\sigma^2}{|X_r|^2} & 0 \\ 0 \ldots & 0 & \frac{2\sigma^4}{n} \end{Bmatrix}$$

Por lo tanto, en la desigualdad de Rao—Cramer,

$$\mathbf{M}_\theta(\theta^* - \theta)^T(\theta^* - \theta) \geqslant I^{-1}(\theta), \tag{16}$$

para las primeras r componentes de θ^* se alcanza la igualdad. Para la componente $r + 1$, la igualdad no puede alcanzarse (aunque asintóticamente, para $n \to \infty$, ambos miembros de (16) se comportan con igualdad), ya que la condición necesaria y suficiente del teorema 2.16.1A aquí no se cumple.

Observación 3. La suposición acerca de la normalidad de ε_i se vuelve poco importante para las afirmaciones (10)—(12), si n es grande (en (11) es mejor realizar la normalización y afirmar la proximidad a la ley normal).

Observación 4. El propio término "regresión" se refiere a la distribución conjunta de dos variables aleatorias ξ y η y significa la curva

$$g(x) = \mathbf{M}(\eta/\xi = x)$$

que también se llama regresión de η en ξ. Por ejemplo, si $(\xi, \eta) \in \Phi_{\gamma,\sigma^2}$, $\gamma = (\gamma_1, \gamma_2)$, $\sigma^2 = |\sigma_{ij}|$, $i, j = 1, 2$, entonces, como hemos visto en los capítulos anteriores, $g(x) = \gamma_2 + \frac{\sigma_{12}}{\sigma_{22}}(x - \gamma_1)$. Esta es una regresión lineal elemental.

Observación 5. La suposición $\xi_i \in \Phi_{0,\sigma^2}$ acerca de la igual distribución de ξ_i cuando se conoce σ^2, puede ser debilitada. Podemos considerar que $\xi_i \in \Phi_{0,\sigma_i^2}$, si σ_i son distintas y conocidas. En este caso, designando por σ la matriz diagonal $\sigma = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{pmatrix}$ e introduciendo nuevas variables $\xi' = \xi\sigma^{-1}$, $X' = X\sigma^{-1}$, $Y' = Y\sigma^{-1}$ (así que $\xi_i' = \xi_i/\sigma_i$, $x_i' = x_i/\sigma_i$, $y_i' = y_i/\sigma_i$), llegaremos al problema de regresión

$$Y' = \alpha X' + \xi'$$

en el que conocemos el vector de observaciones Y' y el regresor X', $\xi' \in \Phi_{0,E_n}$. Es fácil comprobar (el lector puede hacerlo personalmente) que es válido el siguiente análogo del teorema 1

Teorema 2. *La estimación*

$$\alpha^* = Y\sigma^{-2}X^T(D')^{-1}, \quad D' = X\sigma^{-2}X^T,$$

es la estimación eficiente no desplazada de α,

$$(\alpha^* - \alpha)(D')^{1/2} \in \Phi_{0,E_r},$$

$$|Y' = \alpha^* X'|^2 = \sum_{i=1}^{n} \frac{\left(y_i - \sum\limits_{k=1}^{r} \alpha_k^* x_{ik}\right)^2}{\sigma_i^2} \in \mathbf{H}_{n-r}.$$

Recurramos de nuevo al teorema 1. Las relaciones (10)—(12) establecidas en este teorema permiten construir conjuntos confidenciales tanto para distintas coordenadas de θ como para el vector θ en total. Por ejemplo,

$$\mathbf{P}_\theta\left(\frac{(n-r)(\sigma^2)^*}{h^{(2)}_{\varepsilon/2}} < \sigma^2 < \frac{(n-r)(\sigma^2)^*}{h^{(1)}_{\varepsilon/2}}\right) = 1 - \varepsilon, \tag{17}$$

y si X_k son ortogonales, entonces

$$\mathbf{P}_\theta\left(|\alpha_k - \alpha_k^*| < \frac{t_{\varepsilon/2}\sigma^*}{|X_k|}\right) = 1 - \varepsilon, \tag{18}$$

donde $\mathbf{T}_{n-r}((-t_{\varepsilon/2},\ t_{\varepsilon/2})) = 1 - \varepsilon$, $\mathbf{H}_{n-r}((h^{(1)}_{\varepsilon/2},\ h^{(2)}_{\varepsilon/2})) = 1 - \varepsilon$.

Supongamos que X_k son ortogonales. Designemos por $\overline{\alpha}$ el "subvector" del vector α, definido en el corolario 1. En virtud del teorema 1 es natural construir el conjunto confidencial para $\overline{\alpha}$ valiéndose de la relación

$$\frac{|(\overline{\alpha} - \overline{\alpha}^*)\overline{X}|^2}{(n-r)(\sigma^2)^*} < f_\varepsilon. \tag{19}$$

El valor de f_ε, correspondiente al nivel disponible $1 - \varepsilon$, se determina de manera conocida (véase el capítulo 3), o sea, mediante la distribución de Fisher $\mathbf{F}_{l,n-r}$ con l, $n - r$ grados de libertad.

Si se conoce σ^2, el intervalo confidencial será definido por la relación

$$|(\overline{\alpha} - \overline{\alpha}^*)\overline{X}|^2 < \sigma^2 h_\varepsilon, \tag{20}$$

donde h_ε corresponde a la distribución $\mathbf{H}_l$.

En los problemas de regresión puede resultar que también sea necesario estimar el valor de la superficie de regresión $y = \alpha z^T$ en un nuevo punto dado de antemano, $z = (z_1, \ldots, z_r) \in R^r$. Pongamos $y^* = \alpha^* z^T$. Entonces, como antes, hallamos

$$y^* - y = (\alpha^* - \alpha)z^T = \xi X^T D^{-1} z^T \in \Phi_{0,d^2},$$

$$d^2 = \sigma^2 z D^{-1} z^T, \quad \frac{y^* - y}{d\sigma^*} \in \mathbf{T}_{n-r}.$$

Esto da la posibilidad de construir los intervalos confidenciales para y.

Cabe señalar que la determinación de la región confidencial para la superficie de regresión es "en general" un problema más complejo (compárese con [30]). La población de las superficies que entran en el conjunto confidencial será determinada por el conjunto confidencial para θ construido, por ejemplo con ayuda de (10), (11) (véase el § 3.8). Esto se expone más detalladamente en [30].

3. Verificación de las hipótesis con respecto a la regresión lineal. Aquí toquemos dos tipos de problemas.

1) Supongamos que sabemos que la representación (1), (2) tiene lugar. Se necesita verificar la hipótesis de que θ es igual al valor dado de θ' o que el conjunto de coordenadas $\theta_{k_1}, \ldots, \theta_{k_l}$ es igual al conjunto de $\theta'_{k_1}, \ldots, \theta'_{k_l}$, mientras que las demás coordenadas se desconocen.

El criterio para verificar tales hipótesis ha de construirse con ayuda de los conjuntos confidenciales (17)—(20) (véase el § 3.8). Supongamos, por ejemplo, que se necesita verificar la hipótesis H_1 la independencia de Y respecto a X para una regresión lineal elemental, o sea, la hipótesis $H_1 = \{\alpha_2 = 0\}$. Entonces, de (18) (o de (14)) obtenemos el criterio de nivel $1 - \varepsilon$ que rechaza H_1 si

$$|\alpha_2^*| \geqslant t_{\varepsilon/2}\sigma^*/|X_2|. \tag{21}$$

En el caso general de la regresión (1) con X_k ortogonales, la hipótesis de independencia de Y respecto a X tendrá la forma $H_1 = \{\overline{\alpha} = 0\}$, donde $\overline{\alpha} = (\alpha_2, \ldots, \alpha_r)$, $x_{i1} \equiv 1$, y para su verificación se puede aprovechar el criterio

$$\frac{|\overline{\alpha}^* \ \overline{X}|^2}{(n - r)(\sigma^2)^*} \geqslant f_\varepsilon, \tag{22}$$

donde $\overline{X}$ y f_ε están definidas en (19) para $l = r - 1$.

También se pueden utilizar los enfoques del § 3.15, donde fue examinada la verificación de la pertenencia de la muestra a una subfamilia paramétrica. Entonces llegaremos al criterio de relación de verosimilitud, el cual, desde cierto punto de vista, será semejante a (22). Si se conoce σ^2, entonces, el c.r.v. para verificar $H_1 = \{\overline{\alpha} = 0\}$ tendrá la forma

$$\sigma^{-2}|\overline{\alpha}^* \ \overline{X}|^2 > h_\varepsilon,$$

donde h_ε es la cuantila $\mathbf{H}_{r-1}$ de orden $1 - \varepsilon$. Este criterio será minimax (véase los §§ 3.9 y 3.10) para las alternativas correspondientemente separadas.

2) Verificación de la hipótesis de que en la muestra (X, Y) está presente la propia regresión (1), (2). Por estas palabras entendemos la hipótesis de que para α y σ cualesquiera tiene lugar la representación (1), (2), o sea, para α y σ cualesquiera es válida $\sigma^{-1}(Y - \alpha X) \in \Phi_{0,E_n}$. Este es el proble-

ma de pertenencia de Y a una familia paramétrica. Pero como ya hemos señalado, las observaciones en Y no están igualmente distribuidas. Para reducir el problema al caso de distribuciones igualmente distribuidas (véase el § 3.17), haremos uso de la afirmación siguiente, que completa el teorema 1. Consideraremos que X_k son ortogonales.

Teorema 3. *Sea C cualquier matriz ortogonal de orden $n \times n$ que contiene, en calidad de primeras r columnas, las columnas de la matriz $X^T D^{-1/2}$. Entonces, el vector $\delta = (Y - \alpha^* X)C$ tiene coordenadas independientes que poseen la propiedad $\delta_1 = \ldots = \delta_r = 0$, $\delta_i \in \Phi_{0,\sigma^2}$, $i = r + 1, \ldots, n$.*

Ahora bien, el problema se reduce a la verificación de la hipótesis de pertenencia de la muestra $\delta_{r+1}, \ldots, \delta_n$, de volumen $n - r$, a la familia Φ_{0,σ^2} en términos generales (r observaciones se utilizaron para estimar α). Este problema fue examinado en el § 3.17. Para obtener los valores de δ_i es necesario, basándose en las muestras X e Y, calcular sucesivamente los valores de α^*, $Y - \alpha^* X$ y aplicar a $Y - \alpha^* X$ cualquier transformación C dotada de las propiedades indicadas en el teorema 3. Si se conoce σ, llegaremos al problema de verificación de la hipótesis simple de pertenencia de Φ_{0,σ^2}. No obstante, en este caso, para verificar la hipótesis que nos interesa también se puede utilizar el teorema 1, en virtud del cual

$$(n - r)(\sigma^2)^*/\sigma^2 \in \mathbf{H}_{n-r}.$$

Demostración del teorema 3. Si $Z \perp \mathscr{L}[X]$, entonces, las primeras r coordenadas del vector ZC forman el vector $ZX^T D^{-1/2} = 0$. Como $(Y - \alpha^* X) \perp \mathscr{L}[X]$ y $\delta = (Y - \alpha^* X)C$, de aquí resulta que $\delta_1 = \ldots = \delta_r = 0$. Seguidamente,

$$\delta = (Y - \alpha X)C - (\alpha^* - \alpha)XC = \eta - \bar{\eta} D^{-1/2} XC,$$

donde $\eta = \xi C$, $\bar{\eta} = (\eta_1, \ldots, \eta_r) = (\alpha^* - \alpha)D^{1/2} = \xi X^T D^{-1/2}$ y, por consiguiente, δ es el resultado de la transformación lineal sobre η,

$$|\delta|^2 = |Y - \alpha^* X|^2 = |\xi|^2 - |(\alpha - \alpha^*)X|^2 = \sum_{i=1}^{n} \eta_i^2 - |\bar{\eta}|^2 = \sum_{i=r+1}^{n} \eta_i^2,$$

así que $\sum_{i=r+1}^{n} \delta_i^2 = \sum_{i=r+1}^{n} \eta_i^2$. Esto sólo es posible en el caso cuando $(\delta_{r+1}, \ldots, \delta_n)$ es el resultado del giro del vector $(\eta_{r+1}, \ldots, \eta_n)$, o bien, que es lo mismo, el resultado de la transformación ortogonal sobre $(\eta_{r+1}, \ldots, \eta_n)$. En vista de que $\sigma^{-1} \in \Phi_{0,E_n}$, el teorema queda demostrado.

Ejemplo 1. En este ejemplo describiremos el aspecto matemático de un experimento físico con cuya ayuda fue descubierto el efecto de desintegración del mesón φ en dos mesones π (véase [85]). El resultado obtenido tiene carácter estadístico y en él se utilizó, en esencia, el modelo de regresión.

La investigación se refiere al estudio de la interacción de los electrones (e^-) y los positrones (e^+) en los haces que vienen al encuentro. Si la energía total de estas partículas $2E$ se encuentra en el entorno del punto $2E_0 = 1019{,}6$ MeV (fig.8), entonces, al producirse el "choque", de las mismas, como resultado de la acción mutua se forman (a la par con otras) partículas de dos tipos: mesones φ y mesones π. La probabilidad de surgimiento de pares de mesones π durante la interacción de e^+ y e^- conforme a la energía E, se describe con gran precisión por medio de la función lineal que presentamos en forma de (hipótesis H_1)

$$p_1^{\pi\pi}(E) = a_0 + a_1x_1 \quad x = E - E_0, \tag{23}$$

donde a_0, a_1 se desconocen.

Fue planteada la suposición (hipótesis H_2) de que al desintegrarse los mesones φ generados, también pueden aparecer pares de mesones π. Prácticamente es imposible revelar este efecto de un modo directo, ya que se ha establecido que tal fenónemo, si ocurre, se produce muy raramente: no más de una vez en 10^4 desintegraciones de mesones φ. No obstante, gracias al efecto de interferencia de este canal adicional de engendramiento de mesones π, con el canal principal, la probabilidad de que se produzcan dichas partículas será igual no a (23) sino a

$$p_2^{\pi\pi}(E) = [a_0 + a_1x]\left[1 + \frac{b_0 + b_1x + b_2x^2}{x^2 + d^2}\right] \tag{24}$$

(al igual que en (23), ésta es una aproximación muy exacta de una fórmula más compleja, basada en el hecho de que el intervalo de variación que se examina, o sea, $x = E - E_0$, es pequeño en comparación con E_0). En esta igualdad, los coeficientes b_i, al igual que a_i, se desconocen, pero d se conoce.

Para establecer cuál de las dos relaciones, (23) ó (24), tiene lugar en realidad, se ejecutaron $n = 20$ experimentos con distintos valores de energía $E_1, \ldots, E_{20}$.

Los resultados de los experimentos (véase la tabla 1 y la fig. 8) son

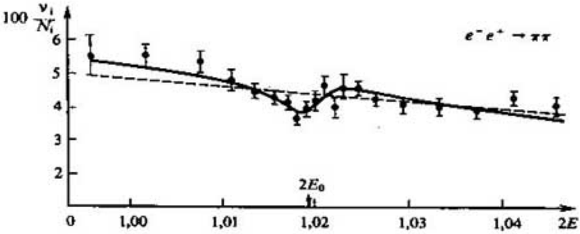

Fig. 8. Las curvas representan las estimaciones de las líneas de regresión para las hipótesis H_1 y H_2.

las cantidades N_i, $i = 1, \ldots, 20$ de interacciones de e^+ y e^-, y las cantidades ν_i de pares de mesones π engendrados con energía E_i. En cada uno de los experimentos efectuados, los números N_i y ν_i son bastante grandes (N_i es del orden de 10^4). En vista de que cuando N_i, es fijo, el número ν_i de pares de mesones π tiene distribución de Bernoulli $\mathbf{B}_{p_i}^{N_i}(p_i = p_1^{\pi\pi}(E_i)$ en la hipótesis H_1, y $p_i = p_2^{\pi\pi}(E_i)$ en la hipótesis H_2), entonces, utilizando la aproximación normal, podemos considerar, con derecho, que tiene lugar la representación

$$y_i \equiv \frac{\nu_i}{N_i} = p_i + \xi_i, \quad \xi_i \in \Phi_{0,\sigma_i^2}$$

Tabla 1. Tabla de los datos experimentales

Número del experimento i	E_i, MeV	N_i	ν_i	Número del experimento i	E_i, MeV	N_i	ν_i
1	497,75	6 960	384	11	510,40	14 322	716
2	500,65	7 908	435	12	510,92	13 470	568
3	503,65	8 102	432	13	511,39	12 008	569
4	505,40	22 259	1080	14	512,17	23 951	1117
5	506,62	16 938	765	15	513,20	27 796	1185
6	507,66	21 728	951	16	514,62	37 771	1539
7	508,40	14 014	603	17	516,58	25 902	1036
8	508,90	13 793	545	18	518,64	27 857	1057
9	509,40	14 075	615	19	520,61	23 228	989
10	509,90	14 867	691	20	522,88	26 482	1066

(en el sumando ξ_i también entran los ruidos eventuales (fondo)). En virtud de (23) y (24) tendremos dos posibles variantes de regresión:

$$p_i = \sum_{k=0}^{1} \alpha_k \psi_k(x_i), \quad \psi_k(x) = x^k, \; k = 0, 1 \tag{25}$$

(hipótesis H_1) y

$$p_i = \sum_{k=0}^{3} \alpha_k \psi_k(x_i), \quad \psi_k(x) = \frac{x^k}{x^2 + d^2}, \quad k = 0, 1, 2, 3 \tag{26}$$

(hipótesis H_2).

Al variar las hipótesis, los valores de σ_i^2 cambian muy poco; éstos pueden ser apreciados muy exactamente y podemos considerar que son conocidos. Entonces, basándose en el teorema 2, la distribución de la estadística

$$\chi^2 = \left| Y' - \sum_k \alpha_k^* \varphi_k \right|^2 = \sum_{i=1}^{n} \left(y_i - \sum_k \alpha_k^* \varphi_k(x_i) \right)^2 \Big/ \sigma_i^2 \tag{27}$$

será $\mathbf{H}_{n-r}$, donde r es el número de parámetros sujetos a estimación α_k ($r = 2$ en la hipótesis H_1, y $r = 4$ en la hipótesis H_2).

Tras realizar los cálculos necesarios conforme a las recomendaciones del teorema 2, obtendremos, para la estadística (27), los valores siguientes: en el primer caso ($r = 2$) $\chi_1^2 = 36{,}8$, y en el segundo ($r = 4$) $\chi_2^2 = 19{,}0$. Los niveles significativos realmente alcanzables (véase el § 3.4) del criterio $\chi^2 > c$ para verificar las hipótesis H_1 y H_2 (como principales) constituirán $H_{18}((0,\ 36{,}8)) = 0{,}9944$ y $H_{16}((0,\ 19{,}0)) = 0{,}731$.

Con otras palabras, la suposición de que falta el canal adicional de engendramiento de pares de mesones π es rechazada por el criterio fundado en la estadística χ^2 con nivel de significación igual, por ejemplo, a 0,99. Al mismo tiempo, la suposición acerca de la existencia de este canal concuerda bien con los resultados experimentales.

Hablando más exactamente, en este problema deberíamos verificar dos hipótesis paramétricas compuestas, correspondientes a las suposiciones (25) y (26) para los valores de las probabilidades de aparición de pares de mesones π. Si utilizamos el criterio de relación de verosimilitud, éste, como es fácil comprobar, se basará en la diferencia de las estadísticas χ^2 correspondientes a los modelos (25) y (26) y, por lo tanto, sus resultados serán aproximadamente los mismos.

4. Estimación y verificación de las hipótesis al existir relaciones lineales. Examinemos, como antes, la regresión lineal (1), (2), pero suponiendo que las coordenadas del vector α están ligadas mediante $s < r$ relaciones lineales

$$\sum_{k=1}^{r} \alpha_k a_{kl} = c_l, \quad l = 1, \ldots, s.$$

En forma matricial estas relaciones pueden escribirse del modo siguiente:

$$\alpha A = c, \tag{28}$$

donde A es una matriz de orden $r \times s$. Supongamos que A es de rango s. En este caso podríamos expresarlas s variables (digamos, $\alpha_{r-s+1}, \ldots$ $\ldots, \alpha_r$) a través de las demás (o sea, a través de $\alpha_1, \ldots, \alpha_{r-s}$), sustituir los valores obtenidos en (1), (2) y volver a obtener el problema estándar de regresión lineal (pero con regresor modificado).

Pero para la exposición ulterior trataremos de resolver este problema de un modo algo distinto. Recurramos a la demostración del teorema 1. El subespacio $\mathscr{A}$ de valores α, definido por las relaciones (28), separa en $\mathscr{L}[X]$ el subespacio de dimensión s y de valores αX, el cual designaremos por $\mathscr{L}_A[X]$. Es evidente que la estimación $\alpha \in \mathscr{A}$ ahora puede efectuarse a base de los mismos procedimientos que hemos utilizado en el teorema 1. La estimación necesaria $\alpha_A^* \in \mathscr{A}$ será determinada, al igual que en el teorema 1, con ayuda de la proyección $\alpha_A^* X$ del vector Y sobre $\mathscr{L}_A[X]$. Ahora bien, a la par con la relación $(Y - \alpha^* X) \perp \mathscr{L}[X]$ tendremos la relación $(Y - \alpha_A^* X) \perp \mathscr{L}_A[X]$ que define unívocamente α_A^*. Para obtener el propio valor de α_A^* es más cómodo hacer uso del enfoque analítico, o sea, aplicar el método de multiplicadores indeterminados de Lagrange para encontrar $\min_\alpha |Y - \alpha X|^2$ a condición de que $\alpha A = c$. Para esto debemos resolver las ecuaciones

$$\alpha A = c, \ \frac{\partial}{\partial \alpha}\left[|Y - \alpha X|^2 + \lambda(\alpha A - c)^T\right] = 0 \qquad (29)$$

(utilizamos los multiplicadores $\lambda_1, \ldots, \lambda_s$ que forman el vector λ y que corresponden a las condiciones (28)). En vista de que $|Y - \alpha X|^2 = (Y - \alpha X)(Y - \alpha X)^T$, la segunda de las ecuaciones (29) adoptará la forma siguiente:

$$-2YX^T - 2\alpha XX^T + \lambda A^T = 0.$$

De aquí hallamos

$$\alpha_A^* = YX^T D^{-1} - \frac{1}{2}\lambda A^T D^{-1} = \alpha^* - \frac{1}{2}\lambda A^T D^{-1}.$$

En virtud de (29), $c = \alpha_A^* A = \alpha^* A - \frac{1}{2}\lambda A^T D^{-1} A$. Como la matriz D está definida positivamente, y el rango de A es s, el rango de la matriz $B = D^{-1/2}A$ también será s, y la matriz $B^T B = A^T D^{-1} A$ también estará positivamente definida (véase el punto 1). Por consiguiente,

$$-\frac{1}{2}\lambda = (c - \alpha^* A)D_A,$$
$$\alpha_A^* = \alpha^* + (c - \alpha^* A)D_A A^T D^{-1}, \qquad (30)$$

donde suponíamos, para abreviar, $D_A = [A^T D^{-1} A]^{-1}$.

El lector puede comprobar que hemos obtenido la e.v.m. del parámetro α a condición de que $\alpha A = c$. Ese mismo resultado (30) también se puede obtener de las consideraciones geométricas, utilizando las relaciones $\alpha_A^* X \in \mathscr{L}_A[X]$ y la ortogonalidad

$$(Y - \alpha_A^* X) \perp \mathscr{L}_A[X],$$
$$(\alpha_A^* - \alpha^*)X = [(Y - \alpha^* X) - (Y - \alpha_A^* X)] \perp \mathscr{L}_A[X]. \qquad (31)$$

Recurramos ahora al problema de *verificación de las hipótesis lineales.* La hipótesis H_1 respecto al parámetro α se llamará hipótesis *lineal* si su forma es $H_1 = \{\alpha A = c\}$, donde las matrices A y c han sido definidas anteriormente.

Inmediatamente podemos señalar que introduciendo el nuevo parámetro $\beta = \alpha A_c$, donde A_c es cualquier matriz no degenerada, cuyas s primeras columnas coinciden con A, reduciremos el problema a la regresión

$$Y = \beta X' + \xi, \quad X' = A_c^{-1} X,$$

y a la verificación de la hipótesis $\bar{\beta} = c$, $\bar{\beta} = (\beta_1, \ldots, \beta_s)$ (véase el punto 2).

También es natural partir de las consideraciones siguientes. Cuanto más se distinga αA de c, tanto más lejos permanecerá αX de $\mathscr{L}_A[X]$ y tanto más se distinguirán los puntos αX y $\alpha^* X$ de $\alpha_A^* Y \in \mathscr{L}_A[X]$. Por eso es natural suponer que la base del criterio para verificar H_1 es la distancia que separa $\alpha_A^* X$ de $\alpha^* X$. Si la hipótesis H_1 es cierta, entonces, en virtud de (31),

$$|(\alpha_A^* - \alpha^*)X|^2 = |Y - \alpha_A^* X|^2 - |Y - \alpha^* X|^2. \tag{33}$$

En virtud de (30) (sustituyendo c por αA), $\alpha_A^* - \alpha^*$ es el resultado de la transformación lineal sobre $\alpha - \alpha^*$. Por eso $(\alpha_A^* - \alpha^*)X$ no depende de $Y - \alpha^* X$ (véase el teorema 1).

Seguidamente, en virtud de (30),

$$\begin{gathered}|(\alpha_A^* - \alpha^*)X|^2 = (\alpha_A^* - \alpha^*)XX^T(\alpha_A^* - \alpha^*)^T = \\ = (c - \alpha^* A)D_A(c - \alpha^* A) = (\alpha^* - \alpha)AD_A A^T(\alpha^* - \alpha)^T.\end{gathered} \tag{34}$$

En vista de que

$$(\alpha^* - \alpha)A = \xi X^T D^{-1} A \in \Phi_{0,\sigma^2 A^T D^{-1} A} = \Phi_{0,\sigma^2 D_A^{-1}},$$

en virtud de (34) y del § 2.2 (punto 4)

$$\frac{1}{\sigma^2}\,|(\alpha_A^* - \alpha^*)X|^2 \in \mathbf{H}_s. \tag{35}$$

De lo dicho y del teorema 1 resulta que

$$\frac{|(\alpha_A^* - \alpha^*)X|^2}{|Y - \alpha^* X|^2} = \frac{|Y - \alpha_A^* X|^2}{|Y - \alpha^* X|^2} - 1 \in \mathbf{F}_{s,n-r}. \tag{36}$$

Las relaciones (35) y (36) nos permiten construir los criterios (basados en la utilización del alejamiento de $\alpha^* X$ respecto a $\alpha_A^* X$) para verificar la hipótesis H_1 en los casos cuando σ^2 se conoce y se desconoce, respectivamente (véase el capítulo 3).

Cabe señalar que H_1 es la hipótesis de pertenencia de α a una subfamilia paramétrica (al existir el parámetro obstaculizador σ^2, si σ^2 se desconoce), y las estadísticas (35) y (36) no son otra cosa sino las estadísticas de la relación de verosimilitud (véanse los §§ 3.10 y 3.15). En efecto, supongamos, por ejemplo, que desconocemos σ^2. Entonces (véanse (5) y (8)),

$$\begin{gathered}\sup_{\alpha,\sigma} f_\theta(Y) = \sup_{\alpha,\sigma}(\sqrt{2\pi}\ \sigma)^{-n}\exp\left\{-\frac{|Y - \alpha X|^2}{2\sigma^2}\right\} = \\ = (\sqrt{2\pi}\ \hat{\sigma}^*)^{-n}\exp\left\{-\frac{|Y - \alpha^* X|^2}{2(\sigma^2)^*}\right\} = \left(\sqrt{2\pi}\ \frac{|Y - \alpha^* X|}{n}\right)^{-n} e^{-n/2}.\end{gathered}$$

El valor de $\sup\limits_{\alpha \in \mathscr{A}, \sigma} f_\theta(Y)$ se calcula exactamente igual. Sólo es preciso señalar que la e.v.m. para α, en el caso de $a\alpha \in \mathscr{A}$, será α_A^*, y la e.v.m. para σ^2 será igual, así como en (8), a $\frac{1}{2}|Y - \alpha_A^* X|^2$. Por eso

$$\sup_{\sigma \in \mathscr{A}_\sigma} f_\theta(Y) = \left(\sqrt{2}\,\frac{|Y - \alpha_A^* X|}{n}\right)^{-n} e^{-n/2},$$

$$\frac{\sup\limits_{\alpha \in \mathscr{A}_\sigma} f_\theta(Y)}{\sup\limits_{\alpha,\sigma} f_\theta(Y)} = \frac{|Y - \alpha^* X|^n}{|Y - \alpha_A^* X|^n}$$

y, por consiguiente, la estadística del criterio de relación de verosimilitud equivale a (36).

Si σ^2 se conoce, como base del criterio para verificar H_1 se puede adoptar la relación (35). Análogamente a lo expuesto más arriba, el lector puede convencerse de que el resultado obtenido también es el criterio de la relación de verosimilitud. Como este criterio es invariable respecto a la sustitución del parámetro (véase el § 3.10), entonces, en virtud de la advertencia y las afirmaciones de los §§ 3.9 y 3.10, se puede afirmar que el c.r.v.

$$|(\alpha_A^* - \alpha^*)X|^2 > \sigma^2 h_\varepsilon,$$

donde h_ε es la cuantila de orden $1 - \varepsilon$ de la distribución $\mathbf{H}_s$, la cual constituirá el criterio minimax de nivel $1 - \varepsilon$ para verificar H_1 frente a las alternativas separadas respectivamente.

Lo dicho más arriba y los resultados de los capítulos 2 y 3 (en particular el § 3.15) dan razones para considerar que los criterios (36), al igual que la estimación (30), también poseen propiedades de optimización. Aquí no nos detendremos más detalladamente en este material. Una exposición más completa de los problemas de regresión se ofrece en [83].

§ 4. Análisis de varianza

Los problemas de análisis de varianza que se exponen en este párrafo pertenecen, en su esencia, a los problemas de regresión. En los últimos de ellos hemos estudiado la dependencia de las observaciones del factor numérico x que podía adoptar cualesquiera valores dados de antemano $x_1, \ldots, x_n$, y a cada uno de ellos le correspondía una sola observación. En los problemas de análisis de varianza suele estudiarse la influencia que ejercen únicamente los factores discretos (uno, dos o más) que pueden tomar exclusivamente un número finito de valores. Para cada uno de estos valores disponemos de un conjunto de observaciones (de una muestra). El análisis de varianza une un grupo de procedimientos estadísticos basados en el análisis de las desviaciones estándar y destinadas a verificar diversas hipótesis y estimar los parámetros relacionados con la influencia de los factores. Los fundamentos del análisis de varianza fueron establecidos por Fisher.

1. Problemas de análisis de varianza como problemas de regresión. El caso de un factor. Supongamos que se dan r muestras independientes

$$Y_1 = (y_{11}, \ldots, y_{1n_1}), \ldots, Y_r = (y_{r1}, \ldots, y_{rn_r})$$

de volúmenes $n_1, \ldots, n_r$ de las poblaciones normales: $Y_k \in \Phi_{\alpha_k,\sigma^2}$. Se supone que las observaciones Y_k, $k = 1, \ldots, r$ se han realizado con diferentes valores de cierto factor cuya importancia nos interesa y que la influencia de este factor se refleja en el valor de la media α_k. Se supone, además, que el valor de la varianza σ^2 es el mismo para todas las muestras y, por regla general, es desconocido. Los problemas de análisis de varianza comprenden la verificación de las hipótesis referentes a los valores $\alpha_1, \ldots, \alpha_r$ y, en particular, de la hipótesis acerca de la homogeneidad de $\alpha_1 = \ldots = \alpha_r = \alpha$ (en el § 1 hemos examinado este último problema), así como las estimaciones de los parámetros α_k y de su variabilidad.

Al igual que los problemas de regresión, el análisis de varianza se aplica ampliamente, sobre todo en la sociología, la agricultura, la biología y la medicina. En calidad de un problema muy típico para aplicar los métodos del análisis de varianza se puede nombrar, por ejemplo, el problema de aclaración de la dependencia que existe entre el contenido de colesterina en la sangre de una persona y su profesión.

Los problemas de análisis de varianza enunciados anteriormente son casos particulares de los problemas de regresión lineal. En efecto, las observaciones y_{ki} pueden representarse en la forma

$$y_{ki} = \alpha_k + \xi_{ki}, \; \xi_{ki} \in \Phi_{0,\sigma^2}, \; k = 1, \ldots, r, \; i = 1, \ldots, n_k. \qquad (1)$$

Formemos el vector

$$Y = ((y_{11}, \ldots, y_{1n_1}; y_{21}, \ldots, y_{2n_2}; \ldots; y_{r1}, \ldots, y_{rn_r})$$

y el vector ξ observando esa misma regla. Entonces, las relaciones (1) pueden ser escritas en la forma matricial $Y = \alpha X + \xi$, donde X es una matriz de dimensión $r \times n$, $n = n_1 + \ldots + n_r$ que tiene la forma siguiente:

$$X = \left(\begin{array}{cccc|cccc|c|cccc}
1 & 1 & \ldots & 1 & 0 & 0 & \ldots & 0 & & 0 & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 & 1 & 1 & \ldots & 1 & & 0 & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & & 0 & 0 & \ldots & 0 \\
\ldots & & & & \ldots & & & & & \ldots & & & \\
0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & & 1 & 1 & \ldots & 1
\end{array}\right).$$

Es evidente que las filas de esta matriz (vectores X_j) son ortogonales. La hipótesis $H_1 = \{\alpha_1 = \alpha_2 = \ldots = \alpha_r\}$ puede escribirse del siguiente modo:

$$\alpha A = 0,$$

donde A es una matriz de dimensión $r \times (r-1)$:

$$A = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 \\ -1 & -1 & \ldots & -1 \end{pmatrix}.$$

Es evidente que el rango de A es $r-1$.

Vemos que la verificación de la hipótesis principal H_1 del análisis de varianza no es otra cosa sino el problema de verificación de la hipótesis lineal para la regresión.

Vamos a aclarar qué son las estimaciones eficientes para α y σ^2 halladas en el teorema 3.1. En nuestro caso $|X_k|^2 = n_k$, la matriz $D = XX^T$ de orden $r \times r$ tiene la forma

$$D = \begin{pmatrix} n_1 & 0 & \ldots & 0 \\ 0 & n_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & n_r \end{pmatrix},$$

$$\alpha_k^* = \frac{(Y, X_k)}{(X_k, X_k)} = \frac{1}{n_k} \sum_{i=1}^{n_k} y_{ki} \equiv \bar{y}_{k.}, \tag{3}$$

$$(n-r)(\sigma^2)^* = |Y - \alpha^* X|^2 = \sum_{k=1}^{r} \sum_{i=1}^{n_k} (y_{ki} - \bar{y}_{k.})^2 \equiv Q_2(Y).$$

En este caso, $\alpha_1^*, \ldots, \alpha_r^*, (\sigma^2)^*$ son independientes. Los intervalos confidenciales para los parámetros α, σ^2, así como sus funciones, se construyen al igual que en el § 3.

Para verificar la hipótesis lineal (2) también debemos calcular la e.v.m. α_A^* al existir la condición (2) (véase el punto 4 del párrafo anterior). Aquí, el método más simple consiste en utilizar el enfoque expuesto al principio del punto 4 del párrafo 3, y en expresar $\alpha_1, \ldots, \alpha_r$ a través de variables independientes. En nuestro caso existe una sola variable independiente: supongamos que ésta sea $\alpha_r = \mu$, y $\alpha_A^* = (\mu^*, \ldots, \mu^*)$ donde μ^* minimiza

$$|Y - (\mu, \ldots, \mu)X|^2 = \sum_{k=1}^{r} \sum_{i=1}^{n_k} (y_{ki} - \mu)^2.$$

Es evidente que

$$\mu^* = \bar{y} \equiv \frac{1}{n} \sum_{k=1}^{r} \sum_{i=1}^{n_k} y_{ki},$$

$$|Y - \alpha_A^* X|^2 = \sum_{k=1}^{r} \sum_{i=1}^{n_k} (y_{ki} - \overline{y})^2 \equiv Q(Y) =$$
$$= \sum_{k=1}^{r} \sum_{i=1}^{n_k} (y_{ki} - \overline{y} + \overline{y}_{k.} - \overline{y}_{k.})^2 =$$
$$= \sum_{k=1}^{r} \sum_{i=1}^{n_k} (y_{ki} - \overline{y}_{k.})^2 + \sum_{k=1}^{r} n_k(\overline{y}_{k.} - \overline{y})^2$$

(la suma de los productos mixtos es igual a cero, puesto que $\sum_{i=1} (y_{ki} - \overline{y}_{k.}) = 0$). Si la hipótesis H_1 es cierta, entonces, en virtud de (3.33), (3) y de la igualdad recién obtenida,

$$|(\alpha_A^* - \alpha^*)X|^2 = Q(Y) - Q_2(Y) = \sum_{k=1}^{r} n_k(\overline{y}_{k.} - \overline{y})^2 \equiv Q_1(Y).$$

En virtud de (3.36), al cumplirse H_1 obtenemos $Q_1(Y)/Q_2(Y) \in \mathbf{F}_{r-1,n-r}$, lo cual no da la posibilidad de construir el criterio $Q_1(Y)/Q_2(Y) > f_\varepsilon$ (f_ε es la cuantila de $\mathbf{F}_{r-1,n-r}$ de orden $1 - \varepsilon$) para verificar H_1, el cual será el c.r.v. Si se conoce σ^2, el c.r.v. tendrá la forma

$$Q_1(Y) > \sigma^2 h_\varepsilon$$

(h_ε es la cuantila de H_{r-1}) y será el criterio minimax para las alternativas separadas respectivamente (véase el § 3.9).

2. Influencia de dos factores. Enfoque elemental. En los problemas de este apartado se investiga la influencia que los factores de dos tipos ejercen sobre los resultados del experimento. Con arreglo, digamos, a la agricultura, esto puede ser el estudio de la influencia que ejerce la composición del suelo (el factor A adopta r valores) y el método de cultivo (el factor B adopta s valores) sobre la calidad de la cosecha.

Aquí las observaciones pueden representarse en la forma

$$y_{kli} = \alpha_{kl} + \xi_{kli}, \quad \xi_{kli} \in \Phi_{0,\sigma^2}, \tag{4}$$
$$k = 1, \ldots, r, \; l = 1, \ldots, s, \; i = 1, \ldots, n_{kl},$$

y el modelo sometido a investigación, en esencia, no se distinguirá en nada del modelo (1) examinado en el punto 1. Por consiguiente, aquí también son aplicables todos los resultados del § 3, pero su aplicación directa es más voluminosa. Ya de por sí es voluminosa la propia presencia de índices triples. Para simplificar algo el problema, pongamos $n_{kl} \equiv 1$; esto nos permitirá eliminar uno de los índices (índice i en (4)). Además, en este apartado proponemos un enfoque elemental algo distinto, que, independientemente de los teoremas del § 3, permitirá obtener las afirmaciones necesarias para la verificación de las hipótesis fundamentales.

Así pues, examinaremos las muestras $Y_{kl} = y_{kl}$ de volumen unitario, de tal modo que el conjunto de datos experimentales Y aquí será la matriz

$r \times s$ de los números y_{kl} que determinan el resultado del experimento bajo la influencia del k-ésimo factor A y el l-ésimo valor del factor B. Esta matriz puede interpretarse como r muestras (filas) de volumen s, correspondientes a distintos valores del factor A, o bien como s muestras (columnas) de volumen r, correspondientes a distintos valores del factor B. De acuerdo con esto, más adelante precisamente tendrá lugar la agrupación de las observaciones. Pongamos

$$\overline{y}_{k} = \frac{1}{s} \sum_{l=1}^{s} y_{kl}, \ \overline{y}_{.l} = \frac{1}{r} \sum_{k=1}^{r} y_{kl}, \ \overline{y} = \frac{1}{rs} \sum_{k,l} y_{kl}.$$

Es válida la identidad

$$Q(Y) \equiv \sum_{k,l} (y_{kl} - \overline{y})^2 = Q_1(Y) + Q_2(Y) + Q_3(Y), \tag{5}$$

donde

$$Q_1(Y) = s \sum_{k} (\overline{y}_{k.} - \overline{y})^2, \ Q_2(Y) = r \sum_{l} (\overline{y}_{.l} - \overline{y})^2,$$

$$Q_3(Y) = \sum_{k,l} (y_{kl} - \overline{y}_{k.} - \overline{y}_{.l} + \overline{y})^2.$$

Supongamos que la influencia ejercida por los factores es aditiva, o sea, existen a_k y b_l tales que

$$\alpha_{kl} = a_k + b_l, \ k = 1, \ldots, r, \ l = 1, \ldots, s. \tag{6}$$

Es evidente que Q_1 determina la variabilidad de los valores a_k (o sea, está relacionada con el factor A), Q_2 determina la variabilidad de b_l (factor B), y Q_3 es una suma que se origina absolutamente por casualidad. También es evidente que

$$Q_i(Y + a) = Q_i(Y), \quad i = 1, 2, 3. \tag{7}$$

Teorema 1. 1)

$$Q_3(Y)/\sigma^2 \in \mathbf{H}_{(r-1)(s-1)}. \tag{8}$$

2) *Si es cierta la hipótesis* $H_A = \{a_1 = \ldots = a_r = a\}$, *entonces* $Q_1(Y)$ *no depende de* $Q_2(Y)$ *y* $Q_3(Y)$, $Q_1(Y)/\sigma^2 \in \mathbf{H}_{r-1}$. *Una afirmación análoga tiene lugar respecto a* Q_2, *y la hipótesis* $H_B = \{b_1 = \ldots = b_s = b\}$.

3) *Si es válida la hipótesis* $H_1 = \{\alpha_{kl} = \alpha\}$, todas las formas cuadráticas Q_1, Q_2 y Q_3 son independientes.

Demostración. Pongamos, sin limitar la generalidad, $\sigma^2 = 1$. Entonces

$$\mathbf{M} y_{kl} y_{ij} = \begin{cases} \alpha_{kl}\alpha_{ij}, & \text{si } (i, j) \neq (k, l), \\ \alpha_{kl}^2 + 1, & \text{si } (i, j) = (k, l). \end{cases}$$

De aquí se deduce que

$$\mathbf{M}\left(\sum_{\mathrm{I}} y_{kl}\right)\left(\sum_{\mathrm{II}} y_{kl}\right) = \left(\sum_{\mathrm{I}} \alpha_{kl}\right)\left(\sum_{\mathrm{II}} \alpha_{kl}\right) + m,$$

donde m es el número de sumandos iguales en las sumas $\sum_{\mathrm{I}}$ y $\sum_{\mathrm{II}}$. Utilizando esta igualdad, ahora es fácil obtener que

$$\mathbf{M}(\mathrm{y}_{k.} - \overline{\mathrm{y}})(\overline{\mathrm{y}}_{.l} - \overline{\mathrm{y}}) = (\alpha_{k.} - \overline{\alpha})(\alpha_{.l} - \overline{\alpha}) = (a_k - \overline{a})(b_l - \overline{b}) \qquad (9)$$

en caso de acuerdos naturales respecto a las designaciones $\alpha_{k.}$, $\alpha_{.l}$, $\overline{\alpha}$, $\overline{a}$, $\overline{b}$. Si es cierta la hipótesis $H_A = \{a_1 = \ldots = a_r = a\}$, la esperanza matemática en (9) es igual a cero. Como en este caso $\mathbf{M}(\overline{\mathrm{y}}_{k.} - \overline{\mathrm{y}}) = \alpha_{k.} - \overline{\alpha} = 0$, el hecho establecido quiere decir que el conjunto de variables aleatorias $\{\overline{\mathrm{y}}_{k.} - \overline{\mathrm{y}}\}$ no depende de $\{\overline{\mathrm{y}}_{.l} - \overline{\mathrm{y}}\}$.

Análogamente establecemos que para cualesquiera k, l, i,

$$\mathbf{M}(\mathrm{y}_{kl} - \overline{\mathrm{y}}_{k.})(\overline{\mathrm{y}}_{i.} - \overline{\mathrm{y}}) = 0.$$

Esto quiere decir que la población $\{\overline{\mathrm{y}}_{k.} - \overline{\mathrm{y}}\}$ tampoco depende de $\{\mathrm{y}_{kl} - \overline{\mathrm{y}}_{k.} - \overline{\mathrm{y}}_{.l} + \overline{\mathrm{y}}\}$. Esto significa, a su vez, que al cumplirse $\mathbf{H}_A$, $Q_1(Y)$ no depende de $Q_2(Y)$ y $Q_3(Y)$. El hecho de que $Q_1(Y) \in \mathbf{H}_{r-1}$, se deduce del lema de Fisher (§ 2.32).

Igualmente sucede cuando se cumple la hipótesis H_B. No obstante, si es válida la hipótesis H_1 (o sea, si son válidas H_A y H_B), es evidente que los tres conjuntos de variables aleatorias mencionadas más arriba serán independientes. Esto significa la independencia de $Q_1(Y)$, $Q_2(Y)$ y $Q_3(Y)$.

Nos queda hallar la distribución $Q_3(Y)$. En vista de que esta distribución no depende de a_k y b_l, podemos considerar que $a_k = b_l = 0$ para todos los k y l y, por consiguiente, se cumple H_1. Entonces, de la definición $Q(Y)$ resulta que $Q(Y) \in \mathbf{H}_{rs-1}$. Además, es válida (5), donde $Q_1(Y) \in \mathbf{H}_{r-1}$ y $Q_2(Y) \in \mathbf{H}_{s-1}$. Nos queda utilizar la independencia $Q_i(Y)$ y el lema 3.1. El teorema está demostrado.

Con arreglo a los problemas del punto 1 también se puede aplicar un enfoque análogo.

Del teorema 1 se deduce la posibilidad de construir los siguientes procedimientos estadísticos:

1) Estimación de los parámetros $a_k - a_i$, $b_l - b_j$, σ^2 (los números a_k y b_l en (6) han sido determinados con una exactitud de hasta el último sumando) con ayuda de las estimaciones $\overline{\mathrm{y}}_{k.} - \overline{\mathrm{y}}_{i.}$, $\overline{\mathrm{y}}_{.l} - \overline{\mathrm{y}}_{.j}$, $(\sigma^2)^* = Q_3(Y)/(r-1)(s-1)$. Como, de hecho, las investigaciones realizadas anteriormente coinciden con lo que hemos hecho en el § 3 y en el punto 1 de este párrafo, las estimaciones mencionadas serán eficientes. Los intervalos confidenciales para σ^2, $a_k - a_i$ pueden ser construidos mediante las rela-

ciones (8),

$$\bar{y}_{k.} - \bar{y}_{i.} - (a_k - a_i) \in \Phi_{0,2\sigma^2/s},$$

$$\frac{\bar{y}_{k.} - y_{i.} - (a_k - a_i)}{\sqrt{\dfrac{2Q_3(Y)}{s(r-1)(s-1)}}} \in \mathbf{T}_{(r-1)(s-1)}$$

(para $b_l - b_j$ todo ocurre análogamente).

2) Verificación de la hipótesis H_A con ayuda del criterio $Q_2/Q_3 > f_\varepsilon$. El nivel del criterio constituirá $1 - \varepsilon$ si f_ε es una cuantila de orden $1 - \varepsilon$ de la distribución $\mathbf{F}_{r-1,(r-1)(s-1)}$.

El criterio para verificar H_B: $Q_2/Q_3 > f_\varepsilon$ tendrá una forma análoga, donde f_ε es una cuantila de orden $1 - \varepsilon$ de la distribución $\mathbf{F}_{s-1,(r-1)(s-1)}$.

3) Verificación de la hipótesis H_1 con ayuda del criterio

$$\frac{Q_1 + Q_2}{Q_3} > f_\varepsilon$$

de nivel $1 - \varepsilon$, donde f_ε es una cuantila de orden $1 - \varepsilon$ de la distribución $\mathbf{F}_{r+s-2,(r-1)(s-1)}$.

Los problemas del análisis de varianza se examinan más detalladamente en [82] y [83].

§ 5. Reconocimiento de imágenes

En este párrafo examinaremos brevemente un grupo de problemas para cuya designación, además del nombre "reconocimiento de imágenes", a veces también se utilizan los términos "clasificación" y "análisis discriminante" *).

En el § 3.1 hemos examinado el siguiente problema de verificación de r hipótesis simples. Se dan las distribuciones $\mathbf{P}_1, \ldots, \mathbf{P}_r$ y la muestra X de volumen n. Es preciso determinar cuál de las hipótesis

$$H_j = \{X \in \mathbf{P}_j\} \tag{1}$$

es cierta.

Sin embargo, en los problemas prácticos, las distribuciones $\mathbf{P}_j$ a menudo se desconocen, y en cuanto a ellas sólo podemos juzgar a partir de las muestras.

Así pues, supongamos que tenemos r muestras $X_l = (x_{l1}, \ldots, x_{ln_l})$, $l = 1, \ldots, r$, de volúmenes $n_1, \ldots, n_r$, respectivamente, que corresponden

---

*) Cabe señalar que los últimos dos términos también se usan para designar otros problemas, por ejemplo, aquellos en los que se conocen las distribuciones $\mathbf{P}_l$ en (1).

a r distribuciones *desconocidas* $\mathbf{P}_1, \ldots, \mathbf{P}_r$, y supongamos, además, que tenemos la muestra X. Es necesario resolver otra vez el mismo problema: determinar, cuál de las hipótesis (1) es cierta. Con otras palabras, es necesario establecer cuál de las muestras $X_1, \ldots, X_r$ es la prolongación de la muestra X. Este es precisamente el problema de reconocimiento de imágenes.

Para simplificar la exposición nos limitaremos a estudiar el caso de $r = 2$.

1. Caso paramétrico. Al principio supongamos que $\mathbf{P}_i$ pertenece a cierta familia paramétrica $\{\mathbf{P}_\theta\}$ que satisface la condición (A_μ), o sea, $X_1 \in \mathbf{P}_{\theta_1}$, $X_2 \in \mathbf{P}_{\theta_2}$, $X \in \mathbf{P}_\theta$ para ciertos $\theta_1 \neq \theta_2$ y $\theta = \theta_1$ o $\theta = \theta_2$. La primera de estas afirmaciones corresponde a la hipótesis $H_1 = \{X \in \mathbf{P}_{\theta_1}\}$, y la segunda, a la hipótesis $H_2 = \{X \in \mathbf{P}_{\theta_2}\}$.

Supongamos seguidamente, también para simplificar la exposición, que los volúmenes n_1, n_2 y n de las muestras son iguales: $n_1 = n_2 = n$.

Examinemos la muestra unida (X_1, X_2, X) y representémosla como una muestra de volumen n formada por las observaciones (x_{1i}, x_{2i}, x_i) y perteneciente a la distribución $\mathbf{P}_{\theta_1} \times \mathbf{P}_{\theta_2} \times \mathbf{P}_\theta$ que tiene una densidad $f_{\theta_1}(x_1)f_{\theta_2}(x_2)f_\theta(x)$ dependiente del parámetro $\overline{\theta} = (\theta_1, \theta_2, \theta)$. Es evidente que la función de verosimilitud de la muestra (X_1, X_2, X) será igual a

$$f_{\overline{\theta}}(X_1, X_2, X) = f_{\theta_1}(X_1)f_{\theta_2}(X_2)f_\theta(X).$$

Hemos llegado al problema de verificación de la hipótesis H_1 acerca de que el parámetro $\overline{\theta}$ se encuentra en la "curva" $\theta = \theta_1$ frente a la hipótesis alternativa H_2 acerca de que $\overline{\theta}$ se encuentra en otra "curva" $\theta = \theta_2$. Este es el problema de verificación de la hipótesis de pertenencia a una subfamilia paramétrica (véase el § 3.15), pero en el caso cuando la hipótesis alternativa significa la pertenencia a otra subfamilia paramétrica. El examen de este problema es análogo al expuesto en el § 3.15, pero en cuanto a su dificultad técnica sale fuera del marco de este manual. Aquí nos limitaremos a describir brevemente, para el caso del parámetro unidimensional θ, la esencia del resultado. Esta esencia es completamente análoga al contenido del § 3.15: si el parámetro θ ha sido localizado, o sea, si los puntos θ_1 y θ_2 están situados en el entorno de cierto punto θ_0, $|\theta_1 - \theta_2| > b/\sqrt{n}$ y si la familia $\{\mathbf{P}_\theta\}$ satisface en el punto θ_0 las condiciones de regularidad (RR), entonces, el criterio de la relación de verosimilitud

$$\frac{\sup\limits_{\theta_1,\theta_2} f_{\theta_1}(X_1)f_{\theta_2}(X_2)f_{\theta_2}(X)}{\sup\limits_{\theta_1,\theta_2} f_{\theta_1}(X_1)f_{\theta_2}(X_2)f_{\theta_1}(X)} > c \tag{2}$$

será, cuando $n \to \infty$, asintóticamente minimax para verificar H_1 frente a H_2.

La limitación $n_1 = n_2 = n$ no tiene importancia. La misma se elimina al igual que en los planteamientos del § 1.

2. Caso general. En el caso general, cuando no hay razones para suponer que X_i están relacionadas con una familia paramétrica, es posible un enfoque general basado en las mismas ideas que hemos utilizado al construir, en el § 2, los criterios de homogeneidad. En este caso el criterio π para verificar H_1 frente a H_2 será una función de tres muestras, así que $\pi = \pi(X_1, X_2, X)$ será la probabilidad de que se acepte H_2 para (X_1, X_2, X) dadas. Al igual que antes, el criterio no randomizado es definido por la región crítica $\Omega \subset \mathscr{X}^{n_1+n_2+n}$ en el espacio de los valores de (X_1, X_2, X). Por nivel de significación del criterio se entiende el número

$$1 - \varepsilon = \inf_{\mathbf{P}_1 \in \mathscr{P}, \mathbf{P}_2 \in \mathscr{P}} \mathbf{P}_1 \times \mathbf{P}_2 \times \mathbf{P}_1((X_1, X_2, X) \notin \Omega),$$

donde $\mathscr{P}$ es la clase de distribuciones admisibles. El valor

$$\beta_\pi \mathbf{P}_1, \mathbf{P}_2 = \mathbf{P}_1 \times \mathbf{P}_2 \times \mathbf{P}_2((X_1, X_2, X) \in \Omega),$$
$$\mathbf{P}_1 \in \mathscr{P}, \mathbf{P}_2 \in \mathscr{P},$$

es la potencia del criterio en el punto $(\mathbf{P}_1, \mathbf{P}_2)$.

El criterio π se llama conciliable cuando $\beta_\pi(\mathbf{P}_1, \mathbf{P}_2) \to 1$ para $n_1 \to \infty$, $n_2 \to \infty$, $n \to \infty$ y para cualesquiera $\mathbf{P}_1 \neq \mathbf{P}_2$, $\mathbf{P}_1 \in \mathscr{P}$, $\mathbf{P}_2 \in \mathscr{P}$.

Como base para construir los criterios conciliables se puede utilizar el hecho bien conocido, acerca de la aproximación de las distribuciones empíricas $\mathbf{P}^*_{X_1}$ y $\mathbf{P}^*_{X_2}$ para las muestras X_1 y X_2 con $\mathbf{P}_1$ y $\mathbf{P}_2$, respectivamente. Si $d(\mathbf{P}, \mathbf{Q})$ es cierta distancia entre las distribuciones, entonces, en el caso de la hipótesis H_2, la distancia $d(\mathbf{P}^*_{X_2}, \mathbf{P}^*_X)$ debe ser menor que $d(\mathbf{P}^*_{X_1}, \mathbf{P}^*_X)$. Por eso, en calidad de criterio se puede utilizar la desigualdad

$$d(\mathbf{P}^*_{X_2}, \mathbf{P}^*_X) - d(\mathbf{P}^*_{X_1}, \mathbf{P}^*_X) < c$$

que al ser cumplida se acepta H_2. El cálculo de tal tipo de criterios (de sus niveles de significación y de su potencia) suele acompañarse de grandes dificultades (comparadlo con el tipo de problemas más simples dados en el § 2).

Utilizando la *agrupación de observaciones,* en el caso general podemos aplicar el criterio asintóticamente óptimo (2). Supongamos que tal agrupación se ha hecho en las regiones $\Delta_1, \ldots, \Delta_m$ y que $(\nu_{i1}, \ldots, \nu_{im})$ y $(\nu_1, \ldots, \nu_m)$ son las frecuencias con que en estas regiones caen las observaciones de las muestras X_i, $i = 1, 2$, y X, respectivamente. Supongamos, además, que $\theta_i = (\theta_{i1}, \ldots, \theta_{im})$ son las probabilidades $(\mathbf{P}_i(\Delta_i), \ldots, \mathbf{P}_i(\Delta_m))$ de caída en las regiones $\Delta_1, \ldots, \Delta_m$ para las distribuciones $\mathbf{P}_i$, $i = 1, 2$. En vista de que para la muestra agrupada X_i, $i = 1, 2$, la función de verosimilitud

$f_{\theta_i}(X_i)$ es igual a $f_{\theta_i}(X_i) = \prod_{k=1}^{m} \theta_{ik}^{\nu_{ik}}$, el criterio (2) tendrá la forma siguiente:

$$\sup_{\theta_2} \sum_{k=1}^{m} (\nu_{2k} + \nu_k)\ln\theta_{2k} + \sup_{\theta_1} \sum_{k=1}^{m} \nu_{1k}\ln\theta_{1k} -$$

$$- \sup_{\theta_1} \sum_{k=1}^{m} (\nu_{1k} + \nu_k)\ln\theta_{1k} - \sup_{\theta_2} \sum_{k=1}^{m} \nu_{2k}\ln\theta_{2k} > \ln c,$$

o bien

$$\sum_{k=1}^{m} (\nu_{2k} + \nu_k) \ln \frac{\nu_{2k} + \nu_k}{n_2 + n} + \sum_{k=1}^{m} \nu_{1k} \ln \frac{\nu_{1k}}{n_1} >$$

$$> \ln c + \sum_{k=1}^{m} (\nu_{1k} + \nu_k) \ln \frac{\nu_{1k} + \nu_k}{n_1 + n} + \sum_{k=1}^{m} \nu_{2k} \ln \frac{\nu_{2k}}{n_2}. \qquad (3)$$

Los planteamientos análogos también pueden efectuarse para $r > 2$.

CAPÍTULO 5

Enfoque de los problemas de la estadística matemática desde el punto de vista de la teoría de los juegos

En los §§ 1—3 se introducen los conceptos de juegos ordinario y estadístico.

En los §§ 4, 5 se examinan los métodos de búsqueda de las decisiones estadísticas óptimas

El material expuesto en los §§ 6—8 está dedicado a la construcción de las reglas de decisión asintóticamente óptimas.

§ 1. Observaciones preliminares

En los capítulos anteriores hemos examinado una gran cantidad de problemas estadísticos diferentes, unidos, todos ellos, por la circunstancia siguiente: el estadista, basándose en datos experimentales, ha de tomar cierta decisión. En la teoría de las estimaciones, tales decisiones pueden tener forma de estimaciones puntuales θ^*, las cuales deben ser adoptadas en calidad de cierto parámetro desconocido θ. En la teoría de verificación de hipótesis estadísticas, las decisiones pueden adoptar forma de afirmaciones que especifican cuáles suposiciones referentes a la naturaleza del objeto sujeto a investigación son ciertas y cuáles son falsas. Dichas decisiones, al ser erróneas, ofrecen pérdidas ulteriores. Por ejemplo, en la estimación de laboratorio (realizada con la ayuda de una muestra), un error en cuanto al contenido de diversos componentes en el mineral, puede provocar la alteración del régimen óptimo de fusión y el empeoramiento de la calidad del metal fundido. Esto significa que experimentaremos pérdidas materiales, las cuales dependerán de la magnitud del desacierto. Un error relacionado con la eficacia de un medicamento que se comprueba en un grupo elegido de enfermos, evidentemente, también puede provocar pérdidas que, para mantener la uniformidad del enfoque, consideraremos que podrán ser calculadas en ciertas unidades. También tomaremos este mismo acuerdo con respecto a otros problemas de estadística en los que las pérdidas no tienen un carácter material claramente expresado.

Lo dicho nos permite destacar, en los problemas de la estadística matemática, los siguientes cuatro elementos comunes que, de hecho, determinan la esencia de cada problema concreto. Para simplificar la exposición, en lo sucesivo hablaremos exclusivamente de los *problemas de una sola muestra X de volumen fijo n.*

1) Conjunto Θ cuyos elementos $\theta \in \Theta$ determinan el estado del objeto sujeto a investigación. Si se conoce θ no habrá necesidad de construir una decisión estadística. El conjunto Θ también se denomina conjunto de parámetros, aunque θ también pueden admitir una interpretación más amplia (por ejemplo, el conjunto Θ puede ser muy rico y coincidir con el conjunto de todas las distribuciones en cierto espacio $\mathscr{X}$).

2) Para obtener alguna información acerca de θ desconocido, el estadista hace un experimento y realiza observaciones respecto a cierta variable aleatoria cuya distribución depende de θ. Con otras palabras, el estadista dispone de la muestra X de la distribución $\mathbf{P}_\theta$. Como ya sabemos, de dicha muestra se puede extraer la información acerca de $\mathbf{P}_\theta$ y, por consiguiente, acerca de θ. Podemos considerar que se cumple la condición (A_0) (véase el § 2.6) en cuanto a la correspondencia biunívoca entre θ y $\mathbf{P}_\theta$.

3) En los problemas de estadística siempre está determinado el conjunto $D = \{\delta\}$ de decisiones que puede tomar el estadista. En la teoría de estimación, el conjunto D suele coincidir con Θ, pero en los problemas de verificación de hipótesis, el conjunto D es finito y el número de sus elementos equivale a la cantidad de hipótesis que se verifican. Si se conoce θ, la decisión $\delta = \alpha(\theta)$ se determina unívocamente. Si se desconoce θ, la decisión δ ha de ser óptima en cierto sentido. Pero la optimización de las decisiones requiere que tengamos la posibilidad de compararlas. Para esto estimaremos que se ha dado la función de pérdidas que determina cuantitativamente la consecuencia de la toma de decisiones.

4) La función de pérdidas $w(\delta, \theta)$ está definida en $D \times \Theta$ e indica las pérdidas que sufriremos si tomamos la decisión δ, en tanto que el objeto sujeto a investigación, al que se refiere la decisión, se halla en estado θ. Consideraremos que $w(\delta, \theta) > 0$ cuando $\delta \neq \varphi(\theta)$, $w(\varphi(\theta), \theta) = 0$.

Si de los cuatro elementos mencionados retiramos el punto 2) acerca de los datos experimentales, obtendremos el objeto que constituye un *juego ordinario de dos personas*, juego en el que el estadista (investigador) desempeña el papel del primer jugador, y la naturaleza, el papel del segundo jugador.

§ 2. Principales conceptos y teoremas relacionados con el juego de dos personas

1. Juego de dos personas.

Definición 1. Llámase *juego de dos personas* la terna (D, Θ, w) compuesta por los conjuntos D y Θ y por la función w que aplica $D \times \Theta$ en la semirrecta $[0, \infty)$. Los elementos δ del conjunto D se denominan *estrategias* (operaciones) *del jugador* I, los elementos $\theta \in \Theta$ se llaman *estrategias del jugador* II, y w es la *función de pérdidas del jugador* I (o la función de ganancia del segundo jugador) que determina las pérdidas $w(\delta, \theta)$ que sufrirá el jugador I si elige la estrategia δ, y las pérdidas que sufrirá el jugador II si elige la estrategia θ.

El principal objetivo de la teoría de los juegos de dos personas consiste en elegir la estrategia óptima del jugador 1 que a menudo identificaremos con nosotros. Para esto es necesario ordenar de algún modo el conjunto de estrategias. No es fácil hacerlo, ya que las pérdidas $w(\delta, \theta)$, con cuya ayuda debemos realizar la ordenación, dependen de dos argumentos, así que, para cada θ, la estrategia δ que minimiza $w(\delta, \theta)$ será, hablando en general, su propia estrategia.

Definición 2. Diremos que la estrategia δ_1 *es mejor que* δ_2, si

$$w(\delta_1, \theta) \leqslant w(\delta_2, \theta) \quad \textit{para todos } \theta \in \Theta \tag{1}$$

y si existe por lo menos un valor de $\theta_1 \in \Theta$ para el cual $w(\delta_1, \theta_1) < w(\delta_2, \theta_1)$.

Si sólo se cumple (1), diremos que la estrategia δ_1 *no es peor que* δ_2.

La estrategia δ_0 para la cual

$$w(\delta_0, \theta) \leqslant w(\delta, \theta) \quad \textit{para todos } \delta \text{ y } \theta$$

la llamaremos estrategia *uniformemente óptima* (o uniformemente mejor).

La estrategia uniformemente mejor asegura las pérdidas mínimas para todos θ. No obstante, por regla general, tales estrategias no existen.

Señalaremos los tres enfoques siguientes para investigar las estrategias óptimas del jugador I:

— determinación de las estrategias uniformemente óptimas en las subclases;

— determinación de las estrategias bayesianas y minimax;

— estudio de la población de todas las estrategias no mejorables (de la llamada clase completa de estrategias).

2. Estrategias uniformemente óptimas en las subclases. Con arreglo a los problemas de la estadística matemática se utiliza a menudo el procedimiento siguiente (véase el § 5). De algunas consideraciones no relacionadas directamente con las pérdidas (consideraciones de simetría, naturalidad del procedimiento, simplicidad de los cálculos, etc.) a veces es posible reducir la clase de estrategias sujetas a examen. Si esta reducción es tal que después de ella existe una estrategia uniformemente óptima, entonces, asimismo se resuelve el problema de elección de la estrategia. Este enfoque debe ir acompañado de investigaciones de la cuestión acerca de si hemos perdido o no (tras reducir la clase) la posibilidad de obtener un resultado mucho mejor. Ejemplos de utilización de tal enfoque (aunque referentes a un objeto más complejo: a los juegos estadísticos) serán examinados en los dos párrafos siguientes. El lector ya sabe de ellos por los capítulos 2 y 3 donde hemos examinado las mejores estimaciones (eficaces) en la subclase de estimaciones no desplazadas, así como los criterios uniformemente más potentes en las subclases de todos los criterios invariantes o no desplazados.

3. Estrategias bayesianas. Estas surgen en los casos en que el segundo jugador elige su estrategia al azar, con cierta distribución (conocida o desconocida) en Θ.

Para tener la posibilidad de examinar posteriormente las estrategias "aleatorias", vamos a suponer que en Θ y D están separadas ciertas σ-álgebras naturales de los subconjuntos $\mathscr{F}_\Theta$ y $\mathscr{F}_D$. Entonces, en $(\Theta, \mathscr{F}_\Theta)$ y $(D, \mathscr{F}_D)$ se pueden definir las distribuciones $\mathbf{Q}$ y π, respectivamente, así que $(\Theta, \mathscr{F}_\Theta, \mathbf{Q})$ y $(D, \mathscr{F}_D, \pi)$ serán los espacios probabilísticos.

La designación de las distribuciones π y $\mathbf{Q}$ induce el espacio probabilístico $(D \times \Theta, \mathscr{F}_{D\times\Theta}, \pi \times \mathbf{Q})$, donde $\mathscr{F}_{D\times\Theta}$ es la σ-álgebra engendrada por los productos directos de los conjuntos de $\mathscr{F}_D$ y $\mathscr{F}_\Theta$. La elección de las σ-álgebras de $\mathscr{F}_D$ y $\mathscr{F}_\Theta$ debe ser tal, que se cumplan las dos condiciones siguientes:

a) $\mathscr{F}_D$ y $\mathscr{F}$ *contienen los conjuntos unipuntuales* $\{\delta\}$ y $\{\theta\}$.

b) La función de pérdidas $w(\delta, \theta)$ es medible con respecto a $\mathscr{F}_{D\times\Theta}$.

Definición 3. Las distribuciones π en $(D, \mathscr{F}_D)$ y $\mathbf{Q}$ en $(\Theta, \mathscr{F}_\Theta)$ se llamarán *estrategias mixtas* o *randomizadas* de los jugadores I y II, respectivamente.

La distribución $\mathbf{Q}$ será frecuentemente llamada distribución *a priori*. El sentido de este término debe estar claro de los capítulos 2 y 3. Además, lo aclararemos adicionalmente en el párrafo siguiente. Los conjuntos de todas las estrategias mixtas de los jugadores I y II (o sea, los conjuntos de todas las distribuciones en $(D, \mathscr{F}_D)$ y $(\Theta, \mathscr{F}_\Theta)$ serán designados por $\tilde{D}$ y $\tilde{\Theta}$. En vista de que $\mathscr{F}_D$ y $\mathscr{F}$ contienen conjuntos unipuntuales, entonces $\tilde{D}$ y $\tilde{\Theta}$ contendrán las distribuciones concentradas en un punto y, por consiguiente, podemos considerar que $\tilde{D}$ y $\tilde{\Theta}$ contienen las estrategias δ y θ que llamaremos *estrategias puras*, a fin de tener la posibilidad de separarlas. El acuerdo, según el cual designaremos con los mismos símbolos δ y θ, respectivamente, las distribuciones de $\tilde{D}$ y $\tilde{\Theta}$ concentradas en un mismo punto δ o θ, no provocará equivocaciones de ningún tipo.

Ahora, las pérdidas $\tilde{w}(\pi, \mathbf{Q})$ provocadas por el uso de estrategias mixtas serán definidas por la igualdad

$$\tilde{w}(\pi, \mathbf{Q}) = \mathbf{M}_{\pi\times Q} w(\delta, \theta) = \int w(u, t)\pi(du)\mathbf{Q}(dt). \tag{2}$$

Así pues, a la par con el juego inicial podemos examinar el juego, $(\tilde{D}, \tilde{\Theta}, \tilde{w})$ con la función de pérdidas (2), el cual se llama *promediación* o *randomización del juego* (D, Θ, w).

Según el acuerdo adoptado escribiremos

$$\tilde{w}(\pi_{(\delta)}, \mathbf{Q}) = \tilde{w}(\delta, \mathbf{Q}), \ \tilde{w}(\pi, \mathbf{Q}_{(\theta)}) = \tilde{w}(\pi, \theta),$$
$$\tilde{w}(\delta, \theta) = w(\delta, \theta),$$

si $\pi_{(\delta)}$ y $\mathbf{Q}_{(\delta)}$ son distribuciones concentradas en los puntos δ y θ, respectivamente.

Es evidente que la randomización del juego (D, Θ, w) significará el paso a un juego con conjuntos de estrategias más ricas, respecto al cual el par inicial es un juego "insertado" que se obtiene al examinar exclusivamente las estrategias puras de ambos jugadores. Como veremos más adelante, los problemas de ordenación de las estrategias en los juegos (D, Θ, w) y $(\tilde{D}, \tilde{\Theta}, \tilde{w})$ se hallan íntimamente ligados.

Definición 4. La estrategia $\pi = \pi_Q$, para la cual

$$\tilde{w}(\pi_Q, \mathbf{Q}) = \inf_{\pi} \tilde{w}(\pi, \mathbf{Q}),$$

se denomina *estrategia bayesiana, correspondiente a la distribución a priori* $\mathbf{Q}$.

Así pues, la estrategia bayesiana no es otra cosa sino la mejor estrategia π para $\mathbf{Q}$ dada en un juego promediado.

La estrategia $\delta_Q \in D$, para la cual $\tilde{w}(\delta_Q, \mathbf{Q}) = \inf \tilde{w}(\pi, \mathbf{Q})$, se denomina *estrategia bayesiana pura.*

Teorema 1. *Si para* $\mathbf{Q}$ *dada existe una estrategia bayesiana mixta* π_Q, *entonces también existirá una estrategia bayesiana pura* δ_Q *tal, que*

$$\tilde{w}(\delta_Q, \mathbf{Q}) = \tilde{w}(\pi_Q, \mathbf{Q}).$$

La **demostración** es casi evidente. Designemos $a = \tilde{w}(\pi_Q, \mathbf{Q})$. Está claro que

$$\tilde{w}(\delta, \mathbf{Q}) \geqslant \inf_{\delta} \tilde{w}(\delta, \mathbf{Q}) \geqslant a.$$

Si admitimos que $\tilde{w}(\delta, \mathbf{Q}) > a$ para todas δ, entonces, realizando la mediación respecto a δ con ayuda de π_Q, obtenemos

$$a = \int \tilde{w}(u, \mathbf{Q})\pi_Q(du) > a.$$

Esta contradicción demuestra el teorema. ◁

Ahora bien, si se alcanza $\inf_{\pi} \tilde{w}(\pi, \mathbf{Q})$, esto también se alcanzará en las estrategias puras.

Si no se alcanza $\inf_{\delta} \tilde{w}(\delta, \mathbf{Q})$, entonces no existirán estrategias bayesianas. En este caso resulta útil el concepto de *estrategia ε-bayesiana* que existe siempre y la cual se define como una estrategia δ_Q para la cual

$$\tilde{w}(\delta_Q,\ \mathbf{Q}) \leqslant \inf_{\delta} \tilde{w}(\delta,\ \mathbf{Q}) + \varepsilon \tag{3}$$

para $\varepsilon > 0$ dado. Sin embargo, en lo sucesivo, para simplificar la exposición nos limitaremos a examinar tan sólo los problemas que contienen las estrategias bayesianas.

La cuestión acerca de la utilización práctica de las estrategias bayesianas es bastante delicada. Si la existencia de la distribución a priori se debe a cierto mecanismo físico real, este enfoque será indiscutible. Pero el enfoque bayesiano también puede ser justificado en los casos en que el mismo esté relacionado con la existencia de ciertas ideas, quizás subjetivas y no siempre bastante completas, las cuales, no obstante, no deben ser rechazadas. En el apartado siguiente (punto 4) se ofrece un análisis más detallado del asunto relacionado con la utilización del enfoque bayesiano.

4. Estrategias minimax. Si se carece de una información a priori respecto a θ, al ordenar las estrategias es posible orientarse hacia la "peor" estrategia del adversario. Si eligemos la estrategia δ, las pérdidas máximas constituirán

$$\sup_{\theta} w(\delta,\ \theta) \equiv w(\delta,\ \uparrow). \tag{4}$$

Esta cantidad sólo depende de δ y, al igual que los valores de $w(\delta,\ \mathbf{Q})$, permite ordenar δ.

Definición 5. La estrategia $\bar{\delta}$ se llama *minimax* si

$$w(\bar{\delta},\ \uparrow) = \inf_{\delta} w(\delta,\ \uparrow) \equiv w^*. \tag{5}$$

El término minimax se forma a base de la unión de las denominaciones de las operaciones en el segundo miembro de la relación

$$w(\bar{\delta},\ \uparrow) = \min_{\delta} \max_{\theta} w(\delta,\ \theta).$$

Es evidente que las estrategias minimax, al igual que las bayesianas, pueden, hablando en general, no existir. En este caso, de un modo análogo a (3), se puede introducir el concepto de estrategia ε-minimax. En los planteamientos ulteriores partiremos del hecho de que en (4) y (5) se alcanzan sup e inf.

En vista de que para cualquier θ

$$w(\bar{\delta},\ \theta) \leqslant w(\bar{\delta},\ \uparrow) = w^*,$$

la estrategia minimax $\bar{\delta}$ se caracteriza por el hecho de que asegura las pérdidas del jugador 1 en cantidad no mayor de w^*.

Definición 6. Los valores

$$w^* = \inf_{\delta} w(\delta,\ \uparrow) \quad (w(\delta,\ \uparrow) = \sup_{\theta} w(\delta,\ \theta)),$$

$$w_* = \sup_{\theta} w(\downarrow,\ \delta) \quad (w(\downarrow,\ \theta) = \inf_{\delta} w(\delta,\ \theta))$$

se llaman, respectivamente, precio superior e inferior del juego. Si $w^* = w_*$, se dice que existe el precio del juego, igual al valor común de w^* y w_*.

De lo dicho anteriormente y de las consideraciones de simetría está claro que el jugador II, actuando análogamente al primero y eligiendo su estrategia $\bar{\theta}$ de las mismas consideraciones minimax, siempre puede asegurar para sí una ganancia no menor de w^*. (Tal estrategia $\bar{\theta}$ sería más correcto llamarla estrategia maximin, pero para ella utilizaremos el mismo término: estrategia minimax). Por lo tanto, *si existe precio del juego*, entonces, eligiendo la estrategia minimax $\bar{\delta}$, aseguraremos para nosotros un *resultado inmejorable* desde el punto de vista siguiente: si el adversario elige $\bar{\theta}$, ninguna otra estrategia nos causará pérdidas θ menores de $w_* = w^*$. Es evidente que

$$w(\bar{\delta}, \bar{\theta}) = w^* = w_*.$$

En el caso general siempre $w^* \geqslant w_*$, ya que para todos δ y θ $w(\delta, \uparrow) \geqslant w(\delta, \theta) \geqslant w(\downarrow, \theta)$ y, por consiguiente,

$$w^* = \inf_{\delta} w(\delta, \uparrow) \geqslant \sup_{\theta} w(\downarrow, \theta) = w_*. \tag{6}$$

Si $w^* > w_*$, entonces, la estrategia minimax $\bar{\delta}$ *se puede mejorar* introduciendo las estrategias mixtas. En esto consiste una de las finalidades principales de estas últimas.

Las estrategias minimax para un juego promediado (si ellas existen) las designaremos por $\bar{\pi}$ y $\bar{\mathbf{Q}}$, respectivamente, y pongamos

$$\tilde{w}^* = \inf_{\pi} \sup_{\mathbf{Q}} \tilde{w}(\pi, \mathbf{Q}), \quad \tilde{w}^* = \sup_{\mathbf{Q}} \inf_{\pi} \tilde{w}(\pi, \mathbf{Q}).$$

Mostremos primeramente que, al promediar el juego, los precios superior e inferior de éste se aproximan.

Teorema 2. $w^* \geqslant \tilde{w}^* \geqslant \tilde{w}_* \geqslant \tilde{w}_*$.

La demostración de este teorema, al igual que la del teorema 1, es muy fácil. En vista de que la mediación del juego puede realizarse en dos etapas: primero por el conjunto D y luego por Θ, para la demostración es suficiente examinar tan sólo la promediación parcial $(\tilde{D}, \Theta, \tilde{w})$ del juego (D, Θ, w). Tenemos

$$\tilde{w}^* = \inf_{\pi} \sup_{\theta} \tilde{w}(\pi, \theta) \leqslant \inf_{\delta} \sup_{\theta} w(\delta, \theta) = w^*.$$

Como para todos π,

$$\tilde{w}(\pi, \theta) = \int w(u, \theta)\pi(du) \geqslant \inf_{\delta} w(\delta, \theta) = w(\downarrow, \theta),$$

entonces, $\inf \tilde{w}(\pi, \theta) \geqslant w(\downarrow, \theta)$,

$$\tilde{w}^* = \sup_{\theta} \inf_{\pi} \tilde{w}(\pi, \theta) \geqslant \sup_{\theta} w(\downarrow, \theta) = w_*.$$

La desigualdad $\tilde{w}^* \geqslant \tilde{w}_*$ ha sido demostrada en (6). $\triangleleft$

El hecho fundamental de la teoría de los juegos consiste en el llamado teorema del minimax, el cual afirma que para suposiciones muy amplias, los *juegos promediados tienen un precio de* $\tilde{w}^* = \tilde{w}_*$ *y para ellos existen estrategias minimax.*

Esta afirmación será enunciada más exactamente en el párrafo siguiente, en una situación más general, con arreglo a los juegos estadísticos.

El juego inicial (D, Θ, w), sobre todo en el caso cuando D y Θ son finitos, por regla general no tiene precio.

Ejemplo 1. Examinemos un juego elemental cuando los conjuntos D y Θ son bipuntuales, $D = \{\delta_1, \delta_2\}$, $\Theta = \{\theta_1, \theta_2\}$. Los valores de la función de pérdidas $w(\delta, \theta)$ se definen por la matriz $\| w(\delta_i, \theta_j) \|$, $i, f = 1, 2$, la cual supondremos que es igual a $\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}$. Esto corresponde, por ejemplo, al juego de adivinación, cuando el jugador I debe adivinar en qué mano el jugador II ha escondido una moneda. La adivinación significa una pérdida nula ($w(\delta_1, \theta_1) = w(\delta_2, \theta_2) = 0$), y el error, una pérdida igual a 1 rublo ($w(\delta_1, \theta_2) = w(\delta_2, \theta_1) = 1$). Es evidente que aquí $w(\delta_i, \uparrow) = 1$, $w^* = 1$, $w(\downarrow, \theta_i) = 0$, $w_* = 0$, por consiguiente, el juego no tiene precio, y el jugador 1 no puede garantizar para sí una pérdida inferior a 1 rublo. El propio concepto de estrategia minimax aquí es inútil.

Examinemos ahora la promediación de este juego. Aquí las clases de estrategias $\tilde{D}$ y Θ son la población de todas las distribuciones en un conjuto bipuntual. Es evidente que cada una de las distribuciones en D y Θ se describe por una probabilidad p y q de elegir las estrategias δ_1 y θ_1, respectivamente. Por eso se puede considerar que $\tilde{D} = [0, 1]$, $\tilde{\Theta} = [0, 1]$. Las pérdidas del jugador I en este juego son iguales a

$$\tilde{w}(p, q) = p(1 - q) + q(1 - p) = p + q - 2pq,$$
$$\tilde{w}(p, \uparrow) = \begin{cases} p + 1 - 2p = 1 - p & \text{para } 2p < 1, \\ p & \text{para } 2p \geqslant 1, \end{cases}$$
$$\tilde{w}^* = 1/2.$$

De un modo análogo hallamos que $\tilde{w}_* = 1/2$. Ahora bien, el juego promediado ya tiene precio y el primer jugador, eligiendo δ_1 y δ_2 con probabilidad $p = 1 - p = 1/2$, puede garantizar para sí una pérdida no mayor de 1/2. Esta estrategia no puede ser mejorada, ya que el jugador II puede garantizar para sí esa misma ganancia, eligiendo $q = 1/2$.

Pero si resulta que el juego promediado no tiene precio (lo cual puede tener lugar tan sólo en los juegos de estructura compleja especial), entonces, la promediación reiterada no dará ningunos resultados, ya que esta promediación repetida coincidirá, en esencia, con la promediación ordinaria.

Los enfoques bayesiano y minimax de la resolución de los problemas de juego tienen gran aplicación en la actividad humana cotidiana. El enfoque bayesiano está orientado hacia la existencia de ciertas nociones, aunque sean aproximadas, del comportamiento del segundo jugador. El enfoque minimax está justificado en los casos en que debemos asegurarnos de una gran derrota.

Ejemplo 2. Un estudiante se prepara para el examen. Supongamos que no es un estudiante ideal y que no ha tenido tiempo suficiente para repasar bien todo el material. Además, el objetivo de este estudiante consiste en obtener la mejor nota posible.

En las condiciones descritas, el estudiante sólo puede estudiar perfectamente parte del material. Por eso, para él son posibles por lo menos dos vías: 1) estudiar en sobresaliente tan sólo las partes que, según la información disponible, el examinador pregunta con más frecuencia; 2) estudiar un poco todo el mateiral para asegurarse una nota buena o satisfactoria. La primera variante corresponderá al enfoque bayesiano, y la segunda, al enfoque minimax.

Claro está que la estrategia uniformemente óptima aquí sería estudiar perfectamente todo el material, pero, según la condición del problema, tal estrategia no es posible.

En las situaciones concretas, las estrategias minimax no siempre son racionales.

Ejemplo 3. Supongamos que $\Theta = [0, 1]$ y que el conjunto $D = \{\delta_1, \delta_2\}$ consta de dos elementos. La función de pérdidas se define por las relaciones (fig. 9)

$$w(\delta_1, \theta) = 1,$$
$$w(\delta_1, \theta) = 4(1 + \varepsilon)\theta(1 - \theta).$$

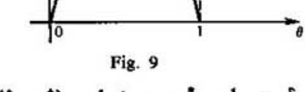

Fig. 9

Aquí $w(\delta_1, \uparrow) = 1$, $w(\delta_1, \uparrow) = 1 + \varepsilon$, $w^* = 1$, y δ_1 será la estrategia minimax, aunque en caso de $\varepsilon > 0$ pequeños, para la "mayoría" de los valores de θ, la estrategia δ_2 será mejor: $w(\delta_2, \theta) < 1$ para θ de la región $\left|\theta - \frac{1}{2}\right| > \frac{1}{2}\sqrt{\frac{\varepsilon}{1 + \varepsilon}}$. Para la "mayoría" de las distribuciones $\mathbf{Q}$ en

$\Theta = [0, 1]$ (cuya masa no está concentrada en el entorno del punto $\theta = 1/2$), las estrategias bayesianas también coincidirán con δ_2.

Los conceptos de estrategia bayesiana y minimax están relacionados entre sí. La siguiente afirmación proporciona el método de averiguación de las estrategias minimax con ayuda de las estrategias bayesianas.

Definición 7. La estrategia $\bar{\pi}$ se llama *igualadora* en el conjunto $\Theta_0 \subset \Theta$ si

1) $\tilde{w}(\bar{\pi}, \theta) = c = \text{const}, \ \theta \in \Theta_0$,

2) $\tilde{w}(\bar{\pi}, \theta) \leqslant c$ para todos θ.

Teorema 3. *Supongamos que existe la distribución a priori* $\bar{\mathbf{Q}}$ *y su estrategia bayesiana correspondiente* $\pi_{\bar{Q}}$, *la cual es igualadora en el portador* $N_{\bar{Q}}$ *de la distribución* $\bar{\mathbf{Q}}$. *Entonces,* $\bar{\pi} = \pi_{\bar{Q}}$ *es una estrategia minimax.*

Si $N_{\bar{Q}} = \Theta$, la estrategia igualadora $\bar{\pi}$ hace "indiferente" el juego del segundo jugador, o sea, lo hace independiente de éste (compárese con el ejemplo 1).

Demostración del teorema 3. Designemos $\sup \tilde{w}(\pi, \theta) = \tilde{w}(\pi, \uparrow)$, $\inf\limits_{\delta} \tilde{w}(\delta, \mathbf{Q}) = \tilde{w}(\downarrow, \mathbf{Q})$. Debemos convencernos de que

$$\tilde{w}(\pi_{\bar{Q}}, \uparrow) = \inf_{\pi} \tilde{w}(\pi, \uparrow).$$

Esto se deduce de las desigualdades siguientes, válidas para cualquier π:

$$\tilde{w}(\pi, \uparrow) \geqslant \tilde{w}(\pi, \bar{\mathbf{Q}}) \geqslant \tilde{w}(\pi_{\bar{Q}}, \bar{\mathbf{Q}}) =$$
$$= \int \tilde{w}(\pi_{\bar{Q}}, t)\bar{\mathbf{Q}}(dt) = c \geqslant \tilde{w}(\pi_{\bar{Q}}, \uparrow). \ \triangleleft$$

A veces es útil la siguiente pequeña generalización del teorema 3.

Teorema 3A. *Supongamos que existen tales sucesiones* $\mathbf{Q}_n$, π_{Q_n} *que* $\tilde{w}(\pi_{Q_n}, \mathbf{Q}_n) \to c$. *Supongamos, además, que existe una estrategia* $\bar{\pi}$ *dotada de la propiedad* $w(\bar{\pi}, \theta) \leqslant c$ *para todos* θ. *Entonces,* $\bar{\pi}$ *es la estrategia minimax.*

La **demostración** es igualmente fácil:

$$\tilde{w}(\pi, \uparrow) \to \tilde{w}(\pi, \mathbf{Q}_n) \geqslant \tilde{w}(\pi_{Q_n}, \mathbf{Q}_n) \to c.$$

Esto puede tener lugar si y sólo si $\inf\limits_{\pi} \tilde{w}(\pi, \uparrow) \geqslant c$. Como $c \geqslant w(\bar{\pi}, \uparrow)$, el teorema queda demostrado.

La distribución $\bar{\mathbf{Q}}$ en el teorema 3, que define la estrategia minimax bayesiana $\pi_{\bar{Q}}$, posee una propiedad magnífica: la misma será la peor en el sentido de que las pérdidas bayesianas $\tilde{w}(\pi_Q, q)$ serán máximas para ella.

Definición 8. La distribución $\bar{\mathbf{Q}}$ se denomina *la menos favorable* o *la peor,* si

$$\tilde{w}(\pi_{\bar{Q}}, \bar{\mathbf{Q}}) = \sup_{\mathbf{Q}} \tilde{w}(\pi_Q, \mathbf{Q}),$$

o, con otras palabras, $\tilde{w}(\downarrow, \bar{\mathbf{Q}}) = \sup\limits_{Q} \tilde{w}(\downarrow, \mathbf{Q})$.

Teorema 4. *Supongamos que el juego* $(\tilde{D}, \tilde{\Theta}, \tilde{w})$ *tiene precio y que ambos jugadores tienen estrategias minimax* $\bar{\pi}$ y $\bar{\mathbf{Q}}$. *Entonces, la distribución* $\bar{\mathbf{Q}}$ *es la peor,* y $\bar{\pi}$ es la estrategia bayesiana $\bar{\pi} = \pi_{\bar{Q}}$ que responde a $\bar{\mathbf{Q}}$.

Observación 1. Del hecho de que, en virtud del teorema 1, a la par con $\pi_{\bar{Q}}$ existe la estrategia bayesiana pura $\delta_{\bar{Q}}$, de ningún modo se deduce que esta última también será minimax.

Observación 2. En virtud del teorema fundamental de los minimax, la condición del teorema 4 acerca de la existencia de precio del juego promediado y de estrategias minimax, no se debe considerar como una limitación considerable.

Necesitaremos la siguiente afirmación auxiliar que enunciaremos en términos del juego inicial (no promediado).

Lema 1. *Supongamos que el juego* (D, Θ, w) *tiene precio y estrategias minimax* $\bar{\delta}$ *y* $\bar{\theta}$ *de ambos jugadores:*

$$w(\bar{\delta}, \uparrow) = \inf_{\delta} w(\delta, \uparrow), \quad w(\downarrow, \bar{\theta}) = \sup_{\theta} w(\downarrow, \theta).$$

Entonces

$$w(\bar{\delta}, \uparrow) = w(\bar{\delta}, \bar{\theta}) = w(\downarrow, \bar{\theta}), \tag{7}$$

$$w^* = w(\bar{\delta}, \bar{\theta}) = w_*. \tag{8}$$

Al contrario, si para ciertos $\bar{\delta}$, $\bar{\theta}$ *se cumple* (7), *entonces es válida* (8), y $\bar{\delta}$, $\bar{\theta}$ *son estrategias minimax.*

Demostración. Para todos δ y θ tenemos

$$w(\delta, \uparrow) \geqslant w(\delta, \theta) \geqslant w(\uparrow, \theta).$$

De aquí resulta

$$w^* = w(\bar{\delta}, \uparrow) \geqslant w(\bar{\delta}, \bar{\theta}) \geqslant w(\downarrow, \bar{\theta}) = w_*. \tag{9}$$

Como, según la condición, $w^* = w_*$, en (9) todos los signos de desigualdad deben sustituirse por signos de igualdad. Esto demuestra (7) y (8).

Al contrario, si es válida (7), entonces

$$w^* = \inf_{\delta} w(\delta, \uparrow) \leqslant w(\bar{\delta}, \uparrow) = w(\downarrow, \bar{\theta}) \leqslant \sup_{\theta} w(\downarrow, \theta) = w_*.$$

En vista de que siempre $w^* \geqslant w_*$, las desigualdades mencionadas significan que $w^* = w_*$ y que las estrategias $\bar{\delta}$ y $\bar{\theta}$ son minimax. El lema queda demostrado.

El punto $(\bar{\delta}, \bar{\theta})$ que posee la propiedad (7) se llama *punto de ensilladura*, el lema 1 se denomina criterio de existencia del punto de ensilladura de las estrategias minimax inmejorables.

Demostración del teorema 4. Apliquemos el lema 1 al juego promediado $(\tilde{D}, \tilde{\Theta}, \tilde{w})$. Entonces obtendremos que

$$\tilde{w}(\bar{\pi}, \bar{\mathbf{Q}}) = \tilde{w}(\downarrow, \bar{\mathbf{Q}}) = \tilde{w}_* = \sup_{Q} \tilde{w}(\downarrow, \mathbf{Q}).$$

De aquí se desprende que la distribución $\bar{\mathbf{Q}}$ es la peor y que $\bar{\pi}$ es la estrategia bayesiana correspondiente a $\bar{\mathbf{Q}}$. El teorema queda demostrado.

El contenido de las afirmaciones citadas anteriormente ahora se puede resumir en forma del criterio siguiente, que tiene carácter minimax y que describe muy ampliamente la relación entre las estrategias minimax y las estrategias bayesianas.

Teorema 5. *Supongamos que el juego* $(\tilde{D}, \tilde{\Theta}, \tilde{w})$ *tiene precio y estrategias minimax. Entonces, las tres condiciones siguientes son equivalentes:*

1) *La estrategia* $\bar{\pi}$ *es minimax.*

2) *La estrategia* $\bar{\pi}$ *es bayesiana e igualadora.*

3) *La estrategia* $\bar{\pi}$ *es bayesiana y corresponde a la peor distribución* $\bar{\mathbf{Q}}$: $\bar{\pi} = \pi_{\bar{\mathbf{Q}}}$.

Demostración. La relación 2) ⇒ 1) se ha demostrado en el teorema 3 (para esto no se necesita la condición del teorema 5). La relación 1) ⇒ 3) se ha establecido en el teorema 4. Necesitamos convencernos de que 3) ⇒ 2), o sea, que la estrategia bayesiana, correspondiente a la peor distribución, es igualadora. Tenemos

$$\tilde{w}_* = \tilde{w}(\bar{\pi}, \bar{\mathbf{Q}}) = \int \tilde{w}(\bar{\pi}, t)\bar{\mathbf{Q}}(dt) \leqslant \sup_t \tilde{w}(\bar{\pi}, t) = \tilde{w}^*.$$

Esto significa que $\int \tilde{w}(\bar{\pi}, t)\bar{\mathbf{Q}}(dt) = \sup_t \tilde{w}(\pi, t)$ y, por consiguiente,

$$\tilde{w}(\bar{\pi}, t) = \tilde{w}(\bar{\pi}, \uparrow) \text{ c.d. } [\bar{\mathbf{Q}}].$$

En vista de que, además, siempre $\tilde{w}(\bar{\pi}, t) \leqslant \tilde{w}(\bar{\pi}, \uparrow)$, entonces $\bar{\pi}$ es una estrategia ingualadora. El teorema queda demostrado.

Volvamos ahora a la cuestión acerca de la aplicación de las clases examinadas de estrategias. Supongamos que no podemos destacar la subclase de estrategias que nos satisfagan, entre las cuales exista la estrategia uniformemente mejor. Supongamos, seguidamente, que disponemos de ciertas nociones acerca del comportamiento del segundo jugador (o sea, de los valores estimados de θ) que, sin embargo, no son suficientes para aplicar el enfoque bayesiano en su forma pura. En estas condiciones el enfoque minimax significará el desprecio de la información que tenemos a nuestra disposición. En tal situación se puede utilizar el enfoque intermedio que consiste en lo siguiente:

1) Primero es necesario protegerse contra las altas pérdidas, o sea, examinar tan sólo las estrategias δ para las cuales $w(\delta, \theta) \leqslant w^* + a$ con valores convenientes de $a > 0$ y para todos θ. El conjunto de estrategias que satisfacen esta desigualdad serán designadas por D_a.

2) En este subconjunto (o sea, en el juego (D_a, Θ, w) ya se puede aplicar el enfoque bayesiano, utilizando las aproximaciones, accesibles a nosotros, para la distribución a priori $\mathbf{Q}$.

Tal enfoque mixto se usa también constantemente en la actividad humana cotidiana. En las condiciones del ejemplo 2 este enfoque significará que el estudiante aprenderá muy superficialmente todo el material (para evitar una nota insatisfactoria) y luego aprenderá mejor lo que se pregunta con más frecuencia.

La utilización matemática del enfoque mixto debe acompañarse de investigaciones de la estabilidad de las pérdidas bayesianas en el juego (D_0, Θ, w) para las variaciones admisibles de $\mathbf{Q}$.

5. Clase completa de estrategias. Si todos los enfoques anteriormente descritos no permiten elegir unívocamente la estrategia, la solución del problema se limite a la descripción de la llamada clase completa de estrategias.

Definición 9. La clase de estrategias $D^o \subset \tilde{D}$ se llama *completa* si para todo $\pi \notin D^c$ existe la estrategia $\pi_0 \in D^c$ que es mejor que π.

La clase D_0^c se denomina *clase completa mínima* si D_0^c es una clase completa, pero a condición de que ninguna de sus propias subclases no sea una clase completa.

Con otras palabras, la clase completa mínima se compone únicamente de estrategias inmejorables.

La utilidad de construcción de la clase completa mínima o de la clase completa, la cual es mucho menor que D, es evidente. Esto da la posibilidad de reducir el juego ($\tilde{D}$, $\tilde{\Theta}$, $\tilde{w}$) al (D^c, $\tilde{\Theta}$, $\tilde{w}$), el cual puede tener una estructura más simple.

El segundo teorema fundamental de la teoría de los juegos consiste en que para amplias suposiciones, *la clase de todas las estrategias bayesianas* $\{\pi_Q\}$, $\mathbf{Q} \in \tilde{\Theta}$, *es una clase completa*. La enunciación exacta de este teorema se dará en el párrafo siguiente. En algunos casos, las clases completas se pueden construir también directamente, utilizando la estructura del juego. Admitamos, por ejemplo, que existe una partición del espacio D en subconjuntos D_b, $D = \bigcup_{b \in B} D_b$, $D_{b_1} \neq D_{b_2}$ cuando $b_1 = b_2$, tal que en cada uno de estos subconjuntos (o sea, para los juegos (D_b, Θ, w)) existe la estrategia uniformemente óptima $\delta_b \in D_b$. Está claro que en este caso la clase $D^c = \{\delta_b\}_{b \in B}$ será completa. Tal enfoque de la construcción de la clase completa será ilustrado en el § 3.

§ 3. Juegos estadísticos

1. Descripción de los juegos estadísticos. Los elementos principales del juego estadístico se forman por la misma terna (D, Θ, w) que hemos examinado en el párrafo precedente. No obstante, se les añade lo siguiente:

1) En los juegos estadísticos el estadista (investigador) desempeña el papel del jugador I, y la naturaleza (más exactamente, la naturaleza del fenó-

meno que se investiga), el papel del jugador II. La naturaleza elige (o "adivina") el parámetro (estrategia) θ que desconocemos y que determina el estado del objeto sometido a investigación. La mayoría de los problemas de la estadística matemática está relacionada, de un modo u otro, con la toma de tales decisiones δ que adivinarían lo más precisamente posible este θ desconocido. En este caso es necesario tener presente que la naturaleza como jugador no tiene por objeto la ganancia máxima (es decir, no intenta causarnos las pérdidas máximas) y desde este punto de vista es un jugador "imparcial" de la elección de sus propias estrategias.

2) En los juegos estadísticos tenemos la posibilidad de "explorar" la estrategia de la naturaleza con ayuda de los experimentos que nos dan en forma de la muestra $X \in \mathbf{P}_\theta$ las indicaciones "sugestivas" de cuál debe ser el valor de θ. Pues, la muestra X de volumen *) n, procedente de la distribución $\mathbf{P}_\theta$ que depende de θ, es un elemento del juego estadístico.

En estas condiciones debemos elegir, evidentemente, nuestra decisión δ en dependencia de X. Por consiguiente, ahora llegan a ser estrategias del estadista todas las funciones $\delta(X)$ que aplican $\mathscr{X}^n$ en D. Estas funciones $\delta(X)$ se llaman *funciones de decisión* o *reglas de decisión*. Nos limitaremos a examinar sólo las funciones $\delta(X)$ que realizan la aplicación medible de $(\mathscr{X}^n, \mathfrak{B}^n_{\mathscr{X}})$ en $(D, \mathfrak{F}_D)$. Designemos por $\mathscr{D}$ el conjunto de todas estas funciones.

El conjunto de estrategias del jugador II (de la naturaleza) Θ queda el anterior.

Si hacemos uso de la decisión $\delta(X)$, y la naturaleza elige θ, nuestras pérdidas constituirán $w(\delta(X), \theta)$. Es una variable aleatoria. Para evitar esta incomodidad, es natural que en calidad de pérdidas para las estrategias $\delta = \delta\,(\cdot) \in \mathscr{D}$ y $\theta \in \Theta$ se tome el valor de la esperanza matemática

$$W(\delta(\cdot),\ \theta) = \mathbf{M}_\theta w(\delta(X),\ \theta) = \int w(\delta(x),\ \theta)\mathbf{P}_\theta(dx), \tag{1}$$

que se llama *función de riesgo* (la aparición de la palabra "riesgo" aquí es natural, ya que la aplicación de $\delta(\cdot)$ da un resultado aleatorio). Si se cumple la condición (A_μ) acerca de la existencia de la densidad $f_\theta(x)$ de la distribución $\mathbf{P}_\theta$ con respecto a cierto ϱ-finita medida μ, entonces la función de riesgo puede escribirse en la forma

$$W(\delta(\cdot),\ \theta) = \int w(\delta(x),\ \theta)\, f_\theta(x)\mu^n(dx).$$

Ahora podemos dar la siguiente

*) En las construcciones de este párrafo podríamos, sin limitar la generalidad, considerar que $n = 1$. Sin embargo, conservaremos el concepto de muestra de volumen n con el fin de dejar válidos los vínculos simples con los resultados de los capítulos precedentes y con las consideraciones posteriores (§§ 6—8).

Una concepción más general de juego estadístico trata de una muestra indefinida $(X_\infty = (x_1, x_2, \ldots)$, en la cual la utilización del elemento x_n va acompañada de las pérdidas $c_n \geqslant 0$ (véase [63]).

Definición 1. Se llama *juego estadístico* la terna $(\mathscr{D}, \Theta, W)$, donde Θ es el conjunto de estrategias de la naturaleza, $\mathscr{D}$ es el conjunto de todas las aplicaciones medibles del espacio $\mathscr{X}^n$ en el conjunto D, y W ha sido definida en (1). Para caracterizar más completamente el juego estadístico, junto con la terna $(\mathscr{D}, \Theta, W)$ se puede considerar también dado el par $(X, \mathbf{P}_\theta)$, donde $X \Subset \mathbf{P}_\theta$.

Ejemplo 1. Supongamos que $\theta \in [0, 1]$ determina el contenido de cierto componente químico de la mena preparada para la fusión. Si tomamos la decisión de que la porción de este componente es igual a $\delta \neq \theta$, y de acuerdo con esta decisión se organiza todo el proceso de fusión, entonces, como resultado, la calidad del metal fundido será peor que cuando $\delta = \theta$, y el consumo de energía será más alto. En otros términos, sufriremos las pérdidas $w(\delta, \theta)$ que serán tanto más grandes cuanto más se distinga δ de θ. Supongamos, para abreviar, que $w(\delta, \theta)$ es proporcional al cuadrado de desviación de δ de θ:

$$w(\delta, \theta) = c(\delta - \theta)^2.$$

(Si la función $w(\delta, \theta)$ es suave y si se examina el entorno de la recta $\delta = \theta$, la suposición simplificadora será aquí únicamente la independencia de c respecto a θ). Como resultado obtendremos el juego (D, Θ, w), en el cual $D = [0, 1]$, $\Theta = [0, 1]$,

$$w(\delta, \uparrow) = \sup_\theta w(\delta, \theta) = \begin{cases} c\delta^2 & \text{para } \delta > 1/2, \\ c(1 - \delta)^2 & \text{para } \delta \leqslant 1/2, \end{cases}$$

$$w^* = \inf_\delta w(\delta, \uparrow) = w(1/2, \uparrow) = c/4.$$

Ahora bien, la estrategia $\delta = 1/2$ es minimax y garantiza las pérdidas $\leqslant c/4$. Como $w_* = 0$, este juego no tiene precio. La randomización del juego no mejora la estrategia minimax $\delta = 1/2$ (da $\tilde{w}_* = c/4$). Le dejamos al lector que él mismo se cerciore de que la estrategia bayesiana δ_Q tiene aquí la forma $\delta_Q = \mathbf{M}_Q\theta = \int t\mathbf{Q}(dt)$ (esto resulta de las igualdades $\tilde{w}(\delta, \mathbf{Q}) = c\mathbf{M}_Q(\delta - \theta)^2 = c\mathbf{M}_Q(\theta - \mathbf{M}_Q\theta)^2 + c\mathbf{M}_Q(\delta - \mathbf{M}_Q\theta)^2$) y que la peor distribución $\bar{\mathbf{Q}}$ tendrá la forma $\bar{\mathbf{Q}}(\{0\}) = \bar{\mathbf{Q}}(\{1\}) = 1/2$. Es evidente que la estrategia bayesiana correspondiente es $\delta_{\bar{Q}} = 1/2$.

Supongamos ahora, que la mena es heterogénea y que tenemos la posibilidad de tomar n pruebas de mineral. Estas pruebas se realizan de modo que los resultados de los análisis de laboratorio para el contenido del componente mencionado en las pruebas sean aleatorios y nos den los valores independientes de $(x_1, \ldots, x_n) = X$ respecto a los cuales se sabe que $\mathbf{M}x_i = \theta$, $\mathbf{D}x_i = b/(\theta)$. En este caso, como decisiones $\delta(X)$ servirán todas las estimaciones posibles $\theta^* = \delta(X)$ del parámetro θ según la muestra X. El riesgo de la función de decisión $\delta(X)$ será igual a

$$W(\delta, \theta) = c\mathbf{M}_\theta(\delta(X) - \theta)^2,$$

y llegamos al problema de determinación de la estimación $\theta^* = \delta(X)$ que minimiza en uno u otro sentido este riesgo. Si ponemos, por ejemplo, $\delta_1(X) = \bar{x}$, obtenemos

$$W(\delta_1, \theta) = \frac{cb(\theta)}{n}. \tag{2}$$

El valor máximo de $b(\theta)$ es igual a $\theta(1-\theta)$ y se alcanza en la distribución x_1 concentrada en los puntos 0 y 1.

Como tal posibilidad se puede excluir, entonces

$$b(\theta) < \theta(1-\theta) \leqslant 1/4, \quad W(\delta_1, \theta) < c/4n.$$

Ahora bien, incluso en el caso de $n = 1$ y cuando se utiliza, quizás, no la mejor estrategia, obtenemos un resultado que es mejor que para la estrategia minimax en el juego sin muestra. La relación (2) también indica que el riesgo converge hacia el cero cuando $n \to \infty$. ◁

De la definición dada anteriormente del juego estadístico se deduce que este último posee un conjunto mucho más rico de estrategias $\mathscr{D}$ en comparación con el juego inicial (D, Θ, w).

Al igual que en el § 2, a la par con el juego $(\mathscr{D}, \Theta, W)$, las estrategias del cual llamaremos *puras*, se pueden examinar *juegos randomizados* o *mixtos* $(\tilde{D}, \tilde{\Theta}, \tilde{W})$. Aquí el conjunto $\tilde{\mathscr{D}}$ es el de las aplicaciones de $\pi(X)$: $\mathscr{X}_n \to \tilde{D}$. Estas aplicaciones deben ser tales que los valores

$$\tilde{w}(\pi(X), \theta) = \int_D w(u, \theta)\pi(X, du)$$

sean variables aleatorias; ($\pi(X, A)$ es la probabilidad del conjunto $A \subset D$ en consonancia con la regla de decisión π). Entonces, por definición, ponemos

$$\tilde{W}(\pi(\cdot), \mathbf{Q}) = \int_\Theta \int_{\mathscr{X}^n} \int_D w(u, t)\pi(x, du)\mathbf{P}_t(dx)\mathbf{Q}(dt).$$

La estrategia $\pi(X)$ se llama *regla randomizada de decisión*.

Las relaciones de orden parcial entre las estrategias, las estrategias uniformemente mejores, bayesianas y minimax, y las clases completas para los juegos estadísticos se definen exactamente igual que para los juegos ordinarios (sustituyendo el conjunto D por $\mathscr{D}$, y las funciones w y $\tilde{w}$, por W y $\tilde{W}$).

Las afirmaciones de los teoremas 2.1—2.5 se extienden por completo a los juegos estadísticos, ya que estas afirmaciones de ningún modo están relacionadas con la naturaleza del conjunto D.

2. Clasificación de los juegos estadísticos. Con la naturaleza de los conjuntos D y Θ está vinculada la siguiente clasificación que separa los tipos principales de los juegos estadísticos:

1) Si $\Theta = A$, $D = A$, donde A es un subconjunto "sólido" en R^k (por ejemplo, un paralelepípedo), $w(t, t) = 0$, $w(t, u) > 0$ para $t \neq u$, obtenemos los problemas de la teoría de estimación puntual del parámetro desconocido θ.

2) Si los conjuntos $\Theta = \{\theta_1, \dots, \theta_r\}$, $D = \{\delta_1, \dots, \delta_r\}$ son finitos y contienen un número igual de elementos, $w(\delta_i, \theta_i) = 0$, $w(\delta_i, \theta_j) > 0$ para $i \neq j$, obtenemos los problemas de verificación de un número finito de hipótesis simples.

3) Si Θ es una región "sólida" en R^k, $D = \{\delta_1, \delta_2\}$ se compone de dos elementos, $w(\delta_1, \theta) = 0$ para $\theta \in \Theta_1$, $w(\delta_2, \theta) = 0$ para $\theta \in \Theta_2$ ($\Theta_1 \cap \Theta_2$ es un vacío) y $w(\delta_i, \theta) > 0$ en los demás casos, llegamos al problema de verificación de las hipótesis $\{\theta \in \Theta_1\}$ y $\{\theta \in \Theta_2\}$.

Son posibles, desde luego, también otras clases de problemas. Hemos destacado estos tres tipos, puesto que han sido examinados en los capítulos 2 y 3. Además, hemos investigado estos problemas partiendo de posiciones puramente "estadísticas", lo que corresponde a una elección especial de las funciones $w(\delta, \theta)$; en el primer grupo de problemas, las pérdidas se han determinado por la desviación estándar, lo que corresponde a la función de pérdidas $w(\delta, \theta) = (\delta - \theta)^2$; en el segundo grupo, las pérdidas se han determinado por la probabilidad de equivocarse, lo que corresponde a la función

$$w(\delta_i, \theta_j) = \begin{cases} 0, & i = j, \\ 1, & i \neq j. \end{cases}$$

Lo mismo se refiere también al tercer grupo de problemas, en el cual hemos utilizado la función de pérdidas

$$w(\delta_1, \theta) = \begin{cases} 0 & \text{para} \quad \theta \in \Theta_1, \\ 1 & \text{para} \quad \theta \in \Theta_2. \end{cases}$$
$$w(\delta_2, \theta) = \begin{cases} 1 & \text{para} \quad \theta \in \Theta_1. \\ 0 & \text{para} \quad \theta \in \Theta_2. \end{cases}$$

Llamaremos funciones *estadísticas* las funciones de pérdidas que corresponden a un enfoque puramente estadístico de los problemas.

La clasificación citada muestra que no existe ningúna diferencia de principio entre los problemas de la teoría de estimación y la verificación de las hipótesis estadísticas. Todo consiste exclusivamente en la naturaleza de los conjuntos Θ y D y en la forma de las funciones de pérdidas.

Tomando como ejemplo esta clasificación, se puede señalar una peculiaridad más de los juegos estadísticos (en adición a los puntos 1 y 2 dados al principio de este párrafo); esta peculiaridad consiste en que en los juegos estadísticos, el conjunto D ora coincide con Θ ora es un conjunto más pobre que Θ.

3. Dos teoremas fundamentales de la teoría de los juegos estadísticos. Vamos a formular ahora los resultados principales de la teoría de los juegos estadísticos. Ya hemos indicado que las afirmaciones de los teoremas 2.1—2.5 quedan válidas, ya que no están relacionadas con la naturaleza de los juegos. Para obtener dos teoremas fundamentales mencionados en el § 2, introduzcamos ciertas suposiciones. No son, ni mucho menos, las suposiciones más generales (de lo contrario, las enunciaciones y demostraciones se complicarían extraordinariamente), pero son bastante amplias para abarcar el grupo más interesante y sustancial de problemas y, en particular, los examinados en los capítulos 2 y 3.

Condición (A). *Cada uno de los conjuntos Θ y D es finito o es un conjunto compacto en R^k.*

Como ya hemos señalado, el caso cuando Θ es finito, y $D \subset R^k$, se puede dejar sin examinar. En los demás tres casos vamos a suponer que la función de pérdidas $w(\delta, \theta)$ satisface la condición siguiente.

Condición (B).

1) *Si $D \subset R^k$, $\Theta \subset R^k$, la función $w(\delta, \theta)$ será continua en $D \times \Theta$.*

2) *Si $\Theta \subset R^k$ y $D = \{\delta_1, \dots, \delta_r\}$ es finito, cada r de las funciones $w(\delta_i, \theta)$, $i = 1, \dots, r$ será continua en Θ.*

Si $\Theta = \{\theta_1, \dots, \theta_r\}$ y $D = \{\delta_1, \dots, \delta_r\}$ son finitas, los valores de $w(\delta_i, \theta_j)$, $i, j = 1, \dots, r$ pueden ser arbitrarios.

Además, exigiremos que se cumpla la

Condición (C). *Disponemos de la muestra $X \in \mathbf{P}_\theta$ de la distribución $\mathbf{P}_\theta$, absolutamente continua para todos θ respecto a cierta medida σ-finita. Si $\Theta \subset R^k$, entonces la densidad $\frac{d\mathbf{P}_\theta}{d\mu}(x) = f_\theta(x)$ es continua en L_1 $(\mathscr{X}, \mathfrak{B}_{\mathscr{X}}, \mu)$ respecto a θ, o sea, para $\theta_m \to \theta$,*

$$\int |f_{\theta_m}(x) - f_\theta(x)| \mu(dx) \to 0. \tag{3}$$

No es difícil comprobar que la continuidad ordinaria $f_\theta(x)$ respecto a θ, para $[\mu]$ c.t. x, contribuye a la continuidad (3).

Teorema 1. *Si se cumplen las condiciones* (A), (B), (C), *el juego promediado* $(\tilde{\mathscr{D}}, \tilde{\Theta}, \tilde{W})$ *tiene precio y estrategias minimax* $\bar{\pi}(X)$ y $\bar{\mathbf{Q}}$:

$$\tilde{W}(\bar{\pi}(\cdot), \uparrow) = \inf_{\pi} \tilde{W}(\pi(\cdot), \uparrow), \quad \tilde{W}(\downarrow, \bar{\mathbf{Q}}) = \sup_{\mathbf{Q}} \tilde{W}(\downarrow, \mathbf{Q}).$$

De los teoremas 2.4 y 2.5 del párrafo precedente sabemos que $\bar{\mathbf{Q}}$ es la peor distribución,

$$\tilde{W}(\pi_{\bar{\mathbf{Q}}}(\cdot), \bar{\mathbf{Q}}) = \sup_{\mathbf{Q}} \tilde{W}(\pi_{\mathbf{Q}}(\cdot), \mathbf{Q}) = \sup_{\mathbf{Q}} \tilde{W}(\downarrow, \mathbf{Q}),$$

y $\bar{\pi}(X) = \pi_{\bar{\mathbf{Q}}}(X)$ es la estrategia bayesiana correspondiente a $\bar{\mathbf{Q}}$.

Sabemos también (véase el teorema 2.5) que para que la estrategia $\bar{\pi}(X)$ sea minimax, es necesario y suficiente que la misma sea bayesiana:

$\overline{\pi}(X) = \pi_Q(X)$ para cierta distribución a priori **Q**, y

$$\tilde{W}(\overline{\pi}(\cdot),\ \overline{\theta} = c = \text{const c.d } [\mathbf{Q}],$$
$$\tilde{W}(\overline{\pi}(\cdot),\ \tilde{\theta}) \leqslant c.$$

Este último criterio del carácter minimax ya fue utilizado reiteradas veces en diferentes situaciones particulares (véanse los §§ 2.11, 3.1, 3.5 y 3.9).

Teorema 2. *Al cumplirse las condiciones* (A), (B), (C), *la clase de todas las estrategias bayesianas será completa.*

En el Suplemento VIII aducimos las demostraciones de los teoremas 1 y 2 en su forma más general, cuando D y Θ son espacios métricos compactos arbitrarios (condición (A)); la función $w(\delta, \theta)$: $D \times \Theta \to R$ es continua respecto a δ y θ en las métricas respectivas (condición (B)); la distribución $\mathbf{P}_\theta$ es continua respecto a θ según la variación (condición (C)).

Las demostraciones de los teoremas 1 y 2 en caso de ciertas suposiciones adicionales, se pueden deducir de [90]. Sin embargo, las demostraciones para el caso de D y Θ finitos se pueden deducir de [7] y [93]. En estas mismas monografías es posible hallar una exposición relativamente completa de los elementos de la teoría general de los juegos estadísticos (y, en particular, la investigación para algunos casos de construcción de la clase completa mínima; véase [93]).

Los teorema 1 y 2 muestran cuán importante es el problema de descripción de la clase de todas las reglas bayesianas de decisión. El siguiente párrafo está dedicado a este problema.

§ 4. Principio bayesiano. Clase completa de funciones de decisión

Hemos visto que por su construcción el juego estadístico es un objeto más complejo que el juego inicial (D, Θ, w). Para este juego, sobre todo si se trata de los conjuntos simples D y Θ (por ejemplo, finitos), la determinación de las estrategias bayesianas y minimax puede ser una tarea relativamente sencilla. Al mismo tiempo, incluso el conjunto D de los juegos estadísticos elementales es de naturaleza muy compleja, y esto puede dificultar considerablemente el estudio de dichos juegos, siempre que los mismos se consideren como juegos ordinarios.

Ejemplo 1. Supongamos que los conjuntos $D = \{\delta_1, \delta_2\}$, $\Theta = \{\theta_1, \theta_2\}$ son bipuntuales, $w(\delta_i, \theta_j) = w_{ij}$, $w_{ii} = 0$, $i, j = 1, 2$. Sea $\mathbf{Q} = (q, 1 - q)$ la distribución a priori en Θ. Entonces,

$$\tilde{w}(\delta_i, \mathbf{Q}) = q w_{i1} + (1 - q) w_{i2}.$$

Por consiguiente, la estrategia bayesiana π_Q tiene la forma

$$\pi_Q(\delta_2) = \begin{cases} 0, & \text{si } \tilde{w}(\delta_1, \mathbf{Q}) < \tilde{w}(\delta_2, \mathbf{Q}) \quad (q w_{21} > (1 - q) w_{12}), \\ 1, & \text{si } \tilde{w}(\delta_2, \mathbf{Q}) \quad \tilde{w}(\delta_1, \mathbf{Q}) \quad (q w_{21} < (1 - q) w_{12}) \end{cases} \tag{1}$$

($\pi_Q(\delta_i)$ es la probabilidad de que se acepte δ_i).

Si

$$\tilde{w}(\delta_1, \mathbf{Q}) = \tilde{w}(\delta_2, \mathbf{Q}) \tag{2}$$

o bien, que es lo mismo, si $q = q = w_{12} + (w_{12} + w_{21})$, entonces, en calidad de π_Q se puede tomar cualquier distribución de π en el conjunto (δ_1, δ_2). De un modo exactamente igual siempre se puede hallar una distribución de $\pi = (p, 1 - p)$ tal, que

$$\tilde{w}(\pi, \theta_1) = \tilde{w}(\pi, \theta_2), \text{ o bien } pw_{12} = (1 - p)w_{21}.$$

La solución de esta ecuación $\bar{p} = w_{21}/(w_{21} + w_{12})$ responde, evidentemente, a la estrategia bayesiana igualadora $\pi_{\bar{Q}}$, $\bar{\mathbf{Q}} = (\bar{q}, 1 - \bar{q})$, la cual, en virtud de los teoremas 2.4 y 2.5, será minimax. La distribución $\bar{\mathbf{Q}}$ será la peor.

Vemos que la "resolución" de este juego se lleva a cabo bastante simplemente. No obstante, si se pasa al juego estadístico, incluso en el caso elemental de $w_{12} = w_{21} = 1$, obtendremos el problema de los criterios bayesianos y minimax para cuya investigación hemos necesitado dos párrafos: 3.1 y 3.2.

Un hecho magnífico, al cual dedicamos el presente párrafo, consiste en que el problema de determinación de las estrategias bayesianas (y, por lo tanto, de la clase completa y de las estrategias minimax) para los juegos estadísticos puede ser reducido, en cierto sentido, al mismo problema para los juegos iniciales (D, Θ, w). Esta reducción se basa en la afirmación siguiente, la cual llamaremos *principio bayesiano*: Sea, como antes,

$$f_\theta(X) = \prod_{i=1}^{n} f_\theta(x_i)$$

la función de verosimilitud de la muestra X y sea ella misma la densidad de X en $\mathscr{X}^n$ respecto a μ^n. Supongamos, además, que la distribución a priori $\mathbf{Q}$ en $(\Theta, \mathfrak{F}_\Theta)$ tiene una densidad $q(t)$ respecto a cierta medida λ (es evidente que esto no es una limitación). Entonces, de acuerdo con el § 2.11, la función $f(x, t) = q(t)f_t(x)$ será la densidad de la distribución compatible de (X, θ) en $\mathscr{X}^n \times \Theta$. Esto quiere decir que la función

$$q(t/x) = \frac{q(t)f_t(x)}{f(x)}, \tag{3}$$
$$f(x) = \int q(t)f_t(x)\lambda(dt),$$

define la densidad condicional de la distribución de θ a condición de que $X = x$. Esta densidad corresponde a la distribución a posteriori $\mathbf{Q}_x$ de la variable aleatoria θ a condición de que $X = x$. La relación (3) se denomina *fórmula de Bayes* (véanse los §§ 2.10 y 2.11).

Teorema 1 (principio bayesiano). *Supongamos que se cumple la condición* (A_μ), *que la distribución a priori en* Θ *tiene una densidad de* $q(t)$, y *que* $\mathbf{Q}_x$, significa la distribución a posteriori de densidad (3), *la cual corresponde a la distribución a priori* $\mathbf{Q}$. *Supongamos, además, que el juego inicial* (D, Θ, w) *para cualquier distribución a priori* $\mathbf{Q}$, *tiene la estrategia bayesiana* π_Q. *Entonces, el juego estadístico* $(\mathscr{D}, \Theta, W)$ *tiene una estrategia bayesiana* $\pi_Q(X)$ *correspondiente a la distribución* $\mathbf{Q}$, *la cual coincide con* π_{Q_x}, *o sea, con la estrategia bayesiana del juego inicial, correspondiente a la distribución a posteriori* $\mathbf{Q}_x$.

La afirmación de este teorema se puede expresar por una sola igualdad

$$\pi_Q(X) = \pi_{Q_x}.$$

Esta reduce el problema planteado, al problema de determinación de la distribución a posteriori $\mathbf{Q}_x$ y al problema de determinación de las estrategias bayesianas para el juego inicial.

El teorema 1 es muy importante para comprender el mecanismo de influencia de la información obtenida de la muestra, sobre la elección de la estrategia óptima. La información a priori, representada por la distribución $\mathbf{Q}$ en Θ, varía continuamente bajo la influencia de los datos experimentales. La estrategia óptima será la que tendrá en cuenta estas variaciones, del modo siguiente: es necesario tomar la estrategia óptima en el juego inicial, pero que ya no corresponde a $\mathbf{Q}$, sino a $\mathbf{Q}_x$.

Demostración del teorema 1. Tenemos

$$\tilde{W}(\pi(\cdot), \mathbf{Q}) = \int_{\Theta}\int_{\mathscr{X}^n} \tilde{w}(\pi(x), t) f_t(x)\mu^n(dx) q(t)\lambda(dt) =$$

$$= \int_{\mathscr{X}^n} f(x)\mu^n(dx) \int_{\Theta} \tilde{w}(\pi(x), t) q(t/x)\lambda(dt). \tag{4}$$

Aquí hemos utilizado (3). El cambio del orden de integración es justo en virtud del carácter no negativo de la función subintegral. La segunda integral en el segundo miembro (4) no es otra cosa sino $\tilde{w}(\lambda(x), \mathbf{Q}_x)$. Pero para cualquier x,

$$\tilde{w}(\pi(x), \mathbf{Q}_x) \geqslant \tilde{w}(\pi_{Q_x}, \mathbf{Q}_x) = \int_{\Theta} \tilde{w}(\pi_{Q_x}, t) q(t+x)\lambda(dt).$$

Sustituyendo esta desigualdad en (4) y volviendo al orden inicial de integración, obtenemos

$$\tilde{W}(\pi(\cdot), Q) \geqslant \int_{\mathscr{X}^n} f(x)\mu^n(dx) \int_{\Theta} \tilde{w}(\pi_{Q_x}, t) q(t+x)\lambda(dt) = \tilde{W}(\pi_Q, \mathbf{Q}).$$

En vista de que aquí $\pi(x)$ es arbitraria, esto quiere decir que

$$\pi_Q(x) = \pi_{Q_x}. \lhd$$

Observación 1. Con fines de precisión, en las consideraciones citadas debemos especificar la mensurabilidad de la función $\tilde{w}(\pi_{Q_x}, t)$ respecto a

$\mathfrak{B}^n \times \mathfrak{F}_\Theta$. Omitimos estas restricciones, ya que éstas tienen un carácter puramente técnico y, al cumplirse las condiciones (A), (B) y (C) del § 3, son completamente innecesarias. El lector puede comprobar personalmente esta última afirmación, utilizando el hecho de que para D y Θ discretos, tal mensurabilidad se establece de un modo evidente, así como el hecho de que el juego arbitrario, al cumplirse las condiciones (A) y (B), puede ser "aproximado" al juego discreto tan exactamente como se quiera.

Volviendo al ejemplo 1, ahora podemos, en virtud del teorema 1, señalar inmediatamente el tipo de estrategias bayesianas para el juego estadístico respectivo. Precisamente de (1) obtenemos

$$\pi_{Q_x}(\delta_2) = \begin{cases} 0, & \text{si} \quad q_X = \dfrac{qf_{\theta_1}(X)}{qf_{\theta_1}(X) + (1-q)f_{\theta_2}(X)} > \dfrac{w_{12}}{w_{12}+w_{21}}, \\ 1, & \text{si} \quad q_X < \dfrac{w_{12}}{w_{12}+w_{21}}. \end{cases} \tag{5}$$

Si

$$q_X = \frac{w_{12}}{w_{12}+w_{21}}, \tag{6}$$

entonces, en calidad de π_{Q_x} se puede tomar cualquier distribución en (δ_1, δ_2). La desigualdad (5) se puede escribir de la forma siguiente:

$$\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)} > \frac{a(1-q)}{q(1-a)}, \quad a = \frac{w_{12}}{w_{12}+w_{21}}. \tag{7}$$

Este es el criterio de relación de verosimilitud que ya conocemos.

Seguidamente,

$$\tilde{W}(\pi_Q, \theta_j) = w_{1j}\mathbf{M}_{\theta_j}\pi_{Q_x}(\delta_1) + w_{2j}\mathbf{M}_{\theta_j}\pi_{Q_x}(\delta_2), \quad j = 1, 2.$$

Supongamos, para abreviar, que la igualdad (6) tiene lugar con $\mathbf{P}_{\theta_j}$ probabilidad de 0, así que la estrategia bayesiana con $\mathbf{P}_{\theta_j}$ probabilidad de 1 será pura, $j = 1, 2$. Entonces,

$$\mathbf{M}_{\theta_j}\pi_{Q_x}(\delta_1) = \mathbf{P}_{\theta_j}\left(\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)} > \frac{a(1-q)}{q(1-a)}\right).$$

$$\tilde{W}(\pi_Q, \theta_1) = w_{21}\mathbf{P}_{\theta_1}\left(\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)} < \frac{a(1-q)}{q(1-a)}\right),$$

$$\tilde{W}(\pi_Q, \theta_2) = w_{12}\mathbf{P}_{\theta_2}\left(\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)} > \frac{a(1-q)}{q(1-a)}\right).$$

De aquí ya no es difícil hallar el valor de $\bar{q}$ correspondiente a la peor distribución $\bar{Q}$, para el cual $\pi_{\bar{Q}_x}$ será la estrategia igualadora, o sea, la estrategia con la que

$$\tilde{W}(\pi_{\bar{Q}}, \theta_1) = \tilde{W}(\pi_{\bar{Q}}, \theta_2).$$

Según los teoremas 2.4 y 2.5, esta estrategia será minimax. Le dejamos al lector que él mismo extienda el procedimiento descrito de determinación de la estrategia minimax, al caso general cuando $\mathbf{P}_{\theta_1}$- o $\mathbf{P}_{\theta_2}$-distribuciones $f_{\theta_1}(X)/f_{\theta_2}(X)$ contienen la componente discreta.

Valiéndonos del teorema 1 podemos, de un modo análogo, obtener la generalización de los resultados de los §§ 3.1 y 3.2 para el caso de D y Θ finitos arbitrarios y de una función arbitraria de pérdidas $w(\delta_i, \theta_j) = w_{ij}$, la cual en este caso también puede llamarse matriz de pérdidas $\|w(\delta_i, \theta_j)\|$. (En los párrafos §§ 3.1 y 3.2 hemos examinado el caso particular de $w_{ij} = 1$ cuando $i \neq j$). Para w_{ij} arbitrarias, la regla bayesiana de decisión tendrá la forma siguiente. Sea $\mathbf{Q} = (q(\theta_1), \dots, q(\theta_r))$, $\mathbf{Q}_x = (q_X(\theta_1), \dots, q_X(\theta_r))$,

$$q_X(\theta_j) = \frac{q(\theta_j) f_{\theta_j}(X)}{\sum_i q(\theta_i) f_{\theta_i}(X)}.$$

Entonces, $\tilde{w}(\delta_i, \mathbf{Q}_X) = \sum_{j=1}^{r} w_{ij} q_X(\theta_j)$ y, por lo tanto,

$\pi_{Q_X}(\delta_k) = 1$, si $\tilde{w}(\delta_k, \mathbf{Q}_X) \leqslant \tilde{w}(\delta_i, \mathbf{Q}_X)$ para todos i, o bien, que es lo mismo, si

$$\sum_{j=1}^{r} w_{kj} f_{\theta_j}(X) q(\theta_j) \leqslant \sum_{j=1}^{r} w_{ij} f_{\theta_j}(X) q(\theta_j).$$

Si existen varios valores de k dotados de esta propiedad (designémoslos por $k_1, \dots, k_s$), entonces, cualquier distribución en $\delta_{k_1}, \dots, \delta_{k_s}$ también será una estrategia bayesiana π_{Q_X}.

La determinación de la estrategia minimax se lleva a cabo del modo siguiente. Supongamos, también para abreviar, que $\mathbf{P}_{\theta_j}$-distribuciones $\tilde{w}(\delta_i, \mathbf{Q}_X)$ no tienen componentes discretas. Entonces,

$$\tilde{W}(\pi_{Q\cdot}, \theta_j) = \sum_{i \neq j} w_{ij} \mathbf{P}_{\theta_j}(\tilde{w}(\delta_i, \mathbf{Q}_X) < \min_{l \neq i} \tilde{w}(\delta_l, \mathbf{Q}_X)).$$

En virtud del teorema 3.1 existe $\bar{\mathbf{Q}} = (\bar{q}(\theta_1), \dots, \bar{q}(\theta_r))$ con la que la estrategia $\pi_{\bar{Q}_X}$ igualará los valores de $\tilde{W}(\pi_{\bar{Q}_X}, \theta_j)$ para todos los valores de j. Esta estrategia será precisamente minimax.

De las consideraciones citadas y del teorema 3.2 también es fácil obtener el tipo de clase completa de estrategias del juego estadístico $(\mathscr{D}, \Theta, \tilde{W})$ en el caso de D y Θ finitos.

Examinemos las estrategias π_{Q_X} que son la distribución aleatoria de tales $\delta_{k_1}, \dots, \delta_{k_s}$ para los cuales

$$\min_i \left(\sum_{j=1}^{r} (w_{k_l, j} - w_{ij}) f_{\theta_j}(X) q(\theta_j) \right) = 0.$$

La clase de tales estrategias (bayesianas), que se obtienen si $q(\theta_1, \ldots, q(\theta_r)$, recorrerán todos los valores posibles y serán una clase completa. Hemos visto que en el caso de $r = 2$ esta clase resulta muy simple y estrecha (véase (7)): consta de las funciones de decisión $\pi(X) = (\pi(X, \delta_1), \pi(X, \delta_2))$, donde $\pi(X, \delta_i)$ son las probabilidades de que se tome la decisión δ_i,

$$\pi(X, \delta_1) = \begin{cases} 1, & \text{si } R(X) > c, \\ p \in [0, 1], & \text{si } R(X) = c, \\ 0 & \text{si } R(X) < c, \end{cases}$$

$$R(X) = \frac{f_{\theta_1}(X)}{f_{\theta_2}(X)}, \quad 0 \leqslant c \leqslant \infty. \tag{8}$$

En los juegos continuos con conjuntos D y Θ para algunas funciones de pérdidas concretas importantes también es posible hallar la forma explícita de las decisiones bayesianas. Supongamos, por ejemplo, que D y Θ son las regiones de R^k, y que la función de pérdidas es cuadrática:

$$w(\delta, \theta) = c\,|\delta - \theta|^2 = c \sum_{i=1}^{k} |\delta_i - \theta_i|^2, \tag{9}$$

donde δ_i, θ_i son las coordenadas δ y θ. Entonces,

$$\tilde{w}(\delta, \mathbf{Q}) = c \int |\delta - t|^2 \mathbf{Q}(dt) = c\mathbf{M}_Q |\delta - \theta|^2.$$

Sabemos que el mínimo de esta expresión se alcanza para $\delta = \mathbf{M}_Q\theta = \int t\mathbf{Q}(dt)$. Esto es, evidentemente, la estrategia bayesiana $\delta_Q = \mathbf{M}_Q\theta$. De aquí y del principio bayesiano resulta que la estrategia bayesiana $\delta_Q(X) = \theta_Q^*$ en el juego estadístico tendrá la forma siguiente:

$$\theta_Q^* = \delta_{Q_X} = \int_{R^k} t\mathbf{Q}_X(dt) = \int_{R^k} tq(t/X)\lambda(dt). \tag{10}$$

Este resultado ya fue obtenido en el capítulo 2.

El riesgo de la estrategia θ_Q^* es igual a $W(\theta_Q^*, \theta) = c\mathbf{M}_\theta |\theta_Q^* - \theta|^2$. La distribución a priori $\overline{\mathbf{Q}}$, para la cual $\mathbf{M}_\theta |\theta_{\overline{Q}}^* - \theta|^2 = \text{const}$, nos ofrecerá la estimación minimax $\overline{\theta}^* = \delta_{\overline{Q}}(X)$. Ejemplos de construcción de estimaciones minimax en esta vía se dan en el § 2.11.

La clase de estimaciones (10), donde $\mathbf{Q}$ recorre los valores en la clase de todas las distribuciones en Θ, es una clase completa.

Examinemos ahora otro caso particular de la función de pérdidas

$$w(\delta, \theta) = c\,|\delta - \theta| \tag{11}$$

y supongamos que $\Theta = R$, $D = R$. Entonces,

$$\tilde{w}(\delta, \mathbf{Q}) = c\mathbf{M}_Q |\delta - \theta| = c \int |\delta - t| \mathbf{Q}(dt) =$$

$$= c \int_{-\infty}^{\delta} (\delta - t)\mathbf{Q}(dt) + c \int_{\delta}^{\infty} (t - \delta)\mathbf{Q}(dt).$$

Utilizando la integración por partes y designando $F(t) = \mathbf{Q}((-\infty, t))$, hallamos

$$\tilde{w}(\delta, \mathbf{Q}) = c \int_{-\infty}^{\delta} (\delta - t)dF(t) - c \int_{\delta}^{\infty} (t - \delta)d(1 - F(t)) =$$

$$= c\left[\int_{-\infty}^{\delta} F(t)dt + \int_{\delta}^{\infty} (1 - F(t))dt \right].$$

La derivada de esta expresión respecto a δ existe c.d. y es igual a $c[2F(\delta) - 1]$. Esta función crece monótonamente y cambia de signo en el punto $\bar{\delta}$, igual a la mediana de la distribución F: $F(\bar{\delta} - 0) \leqslant 1/2$, $F(\bar{\delta} + 0) \geqslant 1/2$. De aquí se deduce que $w(\delta, \mathbf{Q})$ será convexa en cuanto a δ y en el punto $\bar{\delta}$ tendrá el mínimo valor.

En virtud del principio bayesiano esto quiere decir que la *mediana de la distribución a posteriori* $\mathbf{Q}_X$ *será la estimación bayesiana* $\theta^{*}_{Q} = \delta_Q(X)$ *para la distribución a priori* $\mathbf{Q}$ *y la función de pérdidas* (11). Al igual que en el caso (9), esto da la posibilidad de hallar la función de decisión minimax y la clase completa.

De un modo análogo se puede examinar el caso

$$w(\delta, \theta) = c|\delta - \theta|^{\alpha}, \quad \alpha > 0.$$

En conclusión de este párrafo nótese que la función cuadrática de pérdidas (9) en caso de $c = 1$ para los conjuntos continuales D y Θ y la función de pérdidas

$$w(\delta_i, \theta_j) = \begin{cases} 0, & i = j, \\ 1, & i \neq j \end{cases} \tag{12}$$

para D y Θ finitos desempeñan un papel especial en la teoría de los juegos estadísticos. En este caso las funciones de riesgo se convierten en la suma de la varianza y el cuadrado del desplazamiento de la estimación para D y Θ continuales, así como en la probabilidad de equivocarse para D y Θ finitos, respectivamente. Estas características, que son naturales de por sí, nos servían de base para elegir las reglas óptimas en los capítulos 2, 3 y 4. Si un problema estadístico no contiene indicaciones directas concernientes a la forma de la función $w(\delta, \theta)$, entonces con más frecuencia en calidad de $w(\delta, \theta)$ se eligen precisamente estas dos funciones: (9) ó (12). Hemos decidido llamarlas *funciones estadísticas de pérdidas.*

§ 5. Suficiencia, carácter no desplazado e invariación

Los principios de suficiencia, de carácter no desplazado y de invariación sirven para reducir la clase de reglas de decisión. Los mismos consisten en utilizar en calidad de funciones de decisión sólo las reglas de decisión suficientes, no desplazadas e invariantes, respectivamente. La utilización de uno de estos principios, de dos de ellos o de los tres a la vez (si esto es posible) permite, en una serie de casos, reducir hasta tal punto la clase de estrategias sometidas a examen, que su intersección con la clase completa resulta integrada por una sola función de decisión. Esto quiere decir que en la subclase separada existe una estrategia uniformemente mejor (compárese con el punto 1 del § 2) y esto resuelve el problema de elección de la decisión.

Los tres principios son bastante naturales y ya han sido analizados en distintos casos concretos de los capítulos 2 y 3.

El más irrefutable de ellos es el principio de suficiencia, que a menudo no es otra cosa sino el método de descripción de una clase completa.

1. Suficiencia. Supongamos que se cumple la condición ($A\mu$) y que existe la estadística suficiente S, o sea (véase el § 2.12),

$$f_\theta(X) = \psi(\theta, S)\cdot h(X).$$

Supongamos, además, que la distribución a priori $\mathbf{Q}$ tiene una densidad $q(t)$ respecto a cierta medida λ. Entonces, en virtud del principio bayesiano, la estrategia bayesiana será totalmente determinada por la densidad a posteriori

$$q(t/X) = \frac{q(t)f_t(X)}{\int q(u)f_u(X)\lambda(du)} = \frac{q(t)\psi(t, S)}{\int q(u)\psi(u, S)\lambda(du)}, \tag{1}$$

que depende exclusivamente de S. Como cualquier distribución $\mathbf{Q}$ tiene densidad respecto a una medida λ seleccionada respectivamente (se puede poner, por ejemplo, $\lambda = \mathbf{Q}$, $q(t) \equiv 1$), lo dicho significa que todas las reglas bayesianas de decisión $\pi_Q(X)$ serán sólo funciones de S:

$$\pi_Q(X) = p_Q(S).$$

Con otras palabras, cualquier estrategia bayesiana $\pi_Q(X)$ no depende de X al ser fija S.

Ahora supongamos que se cumplen las condiciones (A), (B) y (C) del § 3. Entonces, la afirmación enunciada también atañerá a las estrategias minimax. Esto también significará que todas las reglas de decisión construidas tan sólo como funciones de S (o sea, todas las aplicaciones medibles de Φ / $\tilde{D}$, donde Φ es el espacio en que se hallan los valores de S), forman la clase completa $\mathscr{D}_\Phi$. Esto se deduce del hecho de que $\mathscr{D}_\Phi$ contiene todas las estrategias bayesianas que forman, como sabemos, la clase completa. Evidentemente, la clase $\mathscr{D}_\Phi$ será la mínima para la estadística suficiente mínima S.

Está claro que la clase completa mínima no comprende todas las funciones de S (con valores en $\tilde{D}$), sino tan sólo una parte reducida de las mismas. Eso lo confirma la fórmula (1), de la cual resulta, por ejemplo, que para los conjuntos bipuntuales D y Θ (véase (4.8)), la clase completa está formada por funciones $\pi(X)$ cuya probabilidad $\pi(X, \delta_1)$ de toma de decisión δ_1 tiene forma de indicador del conjunto $\{R(X) > c\}$, donde $R(X) = \psi(\theta_1, S)/\psi(\theta_2, S)$ (véase, para precisar, (4.8)).

Si $D \subset R^k$, $\Theta \subset R^k$ y la función de pérdidas $w(\delta, \theta)$ tiene la forma $w(\delta, \theta) = w(\delta - \theta)$, donde $w(u)$ es una función convexa en R^k, al principio de suficiencia se le puede conferir una forma muy constructiva que permite caracterizar eficientemente la clase completa, o sea, tiene lugar la siguiente generalización del teorema 2.14.1.

Teorema 1 (Blackwell). *Para cualquier función de decisión (estimación) $\theta^* = \delta(X)$ existe la estimación*

$$\theta_S^* = \mathbf{M}_\theta(\theta^*/S)$$

(θ_S^* no depende de θ, ya que S es una estadística suficiente) la cual no es peor que θ^*, o sea, para todos $\theta \in \Theta$,

$$\mathbf{M}_\theta w(\theta_S^* - \theta) \leqslant \mathbf{M}_\theta w(\theta^* - \theta).$$

Demostración. Tiene lugar la siguiente desigualdad de Jensen (véase el § 2.9): si g es una función convexa en R^k; ξ, una variable aleatoria con valores en R^k; y $\mathfrak{F}$, cualquier σ-subálgebra de la σ-álgebra principal, entonces

$$\mathbf{M}(g(\xi)/\mathfrak{F}) \geqslant g(\mathbf{M}(\xi/\mathfrak{F})).$$

Conforme a esta desigualdad,

$$\mathbf{M}_\theta w(\theta^* - \theta) = \mathbf{M}_\theta\{\mathbf{M}_\theta(w(\theta^* - \theta)/S)\} \geqslant$$
$$\geqslant \mathbf{M}_\theta w(\mathbf{M}_\theta(\theta^* - \theta/s)) = \mathbf{M}_\theta w(\theta_S^* - \theta). \triangleleft$$

Si la estadística suficiente S es completa, el teorema 1, junto con el principio de no desplazamiento, permite determinar unívocamente la mejor estimación. En efecto, examinemos la clase K_0 de todas las estimaciones no desplazadas $\theta^* = \delta(X)$:

$$\mathbf{M}_\theta\theta^* = \theta \quad \text{para} \quad \theta^* \in K_0.$$

Entonces, siguiendo exactamente los razonamientos del § 2.14 (teorema 3), nos convencemos de que $\theta_S^* = \mathbf{M}_\theta(\theta^*/S)$ coinciden para todas $\theta^* \in K_0$ y, por consiguiente, la intersección de K_0 y de la clase completa se compone de una sola estimación $\varphi(S)$, la cual es natural llamarla *eficiente.*

De lo dicho se deduce que las *estimaciones eficientes, si existen, serán las mismas para una función convexa arbitraria de pérdidas* $w(\delta - \theta)$. Esto permite utilizar, para cualquiera de estas funciones, todas las afirmaciones de los teoremas respectivos del capítulo 2, obtenidos para $w(u) = u^2$.

Los razonamientos citados ilustran la aplicación compatible de los principios de suficiencia y de carácter no desplazado.

2. Carácter no desplazado. Acabamos de ver qué papel puede desempeñar el principio de carácter no desplazado en la teoría de las estimaciones. En el § 3.6 hemos establecido que un efecto análogo (existencia de criterios no desplazados uniformemente más potentes) puede obtenerse al utilizar los criterios no desplazados en la teoría de verificación de las hipótesis estadísticas.

En el caso general, el carácter no desplazado se define del modo siguiente. Admitamos que el problema de una decisión estadística consiste en "determinar" el valor desconocido de θ y que, por consiguiente, los conjuntos D y Θ coinciden. La función de pérdidas $w(\delta, \theta)$ puede ser arbitraria.

Definición 1. La función de decisión $\delta(X)$ se llama no desplazada si

$$\mathbf{M}_\theta w(\delta(X), \theta) \leqslant \mathbf{M}_\theta w(\delta(X), \theta')$$

para todos θ, $\theta' \neq \theta$.

Con otras palabras, para $v = \theta$ se alcanza $\min\limits_{v} \mathbf{M}_\theta w(\delta(X), v)$. Esto significa que $\delta(X)$, por término medio, se encuentra más cerca de θ desconocido que de cualquier otro punto.

Es fácil notar que la definición de las estimaciones no desplazadas que hemos dado anteriormente es un caso particular de esta afirmación.

Si se verifican dos hipótesis compuestas, $H_1 = \{\theta \in \Theta_1\}$ y $H_2 = \{\theta \in \Theta_2\}$, el conjunto $D = \{\delta_1, \delta_2\}$ puede distinguirse considerablemente de Θ. En este caso, la definición del carácter no desplazado será formalmente algo diferente, aunque su sentido queda invariable, o sea, la definición 1 se puede modificar de tal modo (véase [57]) que la misma pase a la definición siguiente.

Definición 1A. La función de decisión $\delta(X)$ se llama *no desplazada* si

$$\mathbf{M}_\theta w(\delta(X), \theta) \leqslant \mathbf{M}_\theta w(\delta(X), \theta')$$

para todos $\theta \in \Theta_1$, $\theta' \in \Theta_2$ o bien $\theta \in \Theta_2$, $\theta' \in \Theta_1$.

Supongamos, para abreviar, que $w(\delta_1, \theta) = w_1 = \text{const}$ para $\theta \in \Theta_2$; $w(\delta_2; \theta) = w_2 = \text{const}$ para $\theta \in \Theta_1$; $\delta_1 = 0$, $\delta_2 = 1$, y que $\delta(X)$ significa la probabilidad (1 ó 0) de que se acepte H_2. Entonces,

$$\mathbf{M}_\theta w(\delta(X), \theta) = \begin{cases} w_2\mathbf{P}_\theta(\delta(X) = 1) & \text{para} \quad \theta \in \Theta_1, \\ w_1\mathbf{P}_\theta(\delta(X) = 0) & \text{para} \quad \theta \in \Theta_2, \end{cases}$$

$$\mathbf{M}_\theta w(\delta(X), \theta') = \begin{cases} w_1\mathbf{P}_\theta(\delta(X) = 0) & \text{para} \quad \theta \in \Theta_1, \ \theta' \in \theta_2, \\ w_2\mathbf{P}_\theta(\delta(X) = 1) & \text{para} \quad \theta \in \Theta_2, \ \theta' \in \Theta_1 \end{cases}$$

y las desigualdades en la definición 1A quieren decir que

$$w_2\mathbf{P}_{\theta_1}(\delta(X) = 1) \leqslant w_1\mathbf{P}_{\theta_1}(\delta(X) = 0) \quad \text{para} \quad \theta_1 \in \Theta_1,$$
$$w_1\mathbf{P}_{\theta_2}(\delta(X) = 0) \leqslant w_2\mathbf{P}_{\theta_2}(\delta(X) = 1) \quad \text{para} \quad \theta_2 \in \Theta_2,$$

o bien, que es lo mismo,

$$\mathbf{P}_{\theta_1}(\delta(X) = 1) \leqslant \frac{w_1}{w_1 + w_2}, \quad \mathbf{P}_{\theta_2}(\delta(X) = 1) \geqslant \frac{w_1}{w_1 + w_2}.$$

De aquí se deduce que

$$\sup_{\theta \in \Theta_1} \mathbf{M}_\theta \delta(X) \leqslant \inf_{\theta \in \Theta} \mathbf{M}_\theta \delta(X)$$

y que, por consiguiente, el criterio δ no será desplazado desde el punto de vista de las definiciones del § 3.6. Al contrario, si es válida la última desigualdad, el criterio δ no será desplazado desde el punto de vista de la definición 1A al elegir adecuadamente la función de pérdidas $w(\delta, \theta)$, por ejemplo, para $w_1/(w_1 + w_2) = \sup\limits_{\theta \in \Theta_1} \mathbf{M}_\theta \delta(X)$.

Los ejemplos adicionales de utilización del principio no desplazado (además de los resultados obtenidos en el § 3.6) se pueden hallar en [57].

3. Invariación. Hemos visto que la intersección de la clase completa, engendrada por las decisiones "suficientes", con la clase de decisiones no desplazadas puede constar de una sola estrategia. La clase de reglas de decisión invariantes es otra clase natural de estrategias, en la que puede resultar la única decisión inmejorable (compárese con los §§ 2.18, 2.19 y 3.7).

La definición del problema invariante de decisión estadística está relacionada con los grupos de transformaciones en los tres espacios que participan en la definición del juego estadístico: en los espacios D y Θ y en el espacio muestral $\mathscr{X}^n$. La definición se basa en las transformaciones biunívocas medibles g del espacio $\mathscr{X}^n$ en sí, que forman cierto grupo G con la operación de grupo definida como una composición: si $g_1 \in G$ y $g_2 \in G$, entonces $g_2 g_1$ se define como una transformación $x \to g_2(g_1 x)$ que otra vez debe pertenecer a G. Designemos por e la transformación idéntica. Sin embargo, la transformación g^{-1} inversa a g se define como una transformación para la cual $g^{-1}g = e$. La mensurabilidad de $g \in G$ significa que gX, junto con X, será una variable aleatoria en $\mathscr{X}^n$.

Con el grupo introducido G está estrechamente relacionado el concepto de invariación de la familia $\mathbf{P}_\theta$ que hemos definido en los §§ 2.19 y 3.7. Este concepto significa que para $g \in G$ y $\theta \in \Theta$ habrá un elemento $\theta_g \in \Theta$ tal, que

$$\mathbf{P}_\theta(gX \in A) = \mathbf{P}_{\theta_g}(X \in A). \tag{2}$$

Las transformaciones $\bar{g}$ del espacio Θ en sí, definidas por la igualdad $\bar{g}\theta = \theta_g$, al cumplirse la condición (A_0) forman el grupo $\bar{G}$ (véase el § 2.19).

En términos de las esperanzas matemáticas, la condición (2) significa que para cualquier función integrable φ,

$$\mathbf{M}_\theta \varphi(gX) = \mathbf{M}_{\bar{g}\theta} \varphi(X). \tag{3}$$

Definición 2. El problema de decisión estadística, relacionado con el juego estadístico $(\mathscr{D}, \Theta, w)$, $(X, \mathbf{P}_\theta)$, se llama *problema invariante respecto al grupo G*, si la familia $\mathbf{P}_\theta$ es invariante respecto a G, y la función de pérdidas w es invariante respecto a G en el sentido siguiente: para cualesquiera $\delta \in D$, $g \in G$ existirá el único $\delta' \in D$ tal, que

$$w(\delta, \theta) = w(\delta', \bar{g}\theta) \quad \text{para todos} \quad \theta \in \Theta. \tag{4}$$

El valor δ', unívocamente definido respecto a g, lo designaremos por $g'\delta$.

Lema 1. *Las transformaciones g' del espacio D en sí, engendradas por el grupo G, forman el grupo G'.*

Demostración. Mostraremos que la población G' de todas las transformaciones g' está cerrada respecto a la composición y que además es válida la igualdad $g_2' g_1' = (g_2 g_1)'$.

En efecto,

$$w(\delta, \theta) = w(g_1' \delta, \bar{g}_1\theta) = w(g_2' g_1' \delta, \bar{g}_2\bar{g}_1\theta) = w((g_2g_1)'\theta, \overline{(g_2g_1)}\theta).$$

Como $\overline{(g_2g_1)} = \bar{g}_2\bar{g}_1$, entonces, en virtud de la unicidad, $(g_2g_1)' = g_2' g_1'$. El lema queda demostrado.

Así pues, con el principal grupo G de las transformaciones g del espacio $\mathscr{X}^n$ en sí, están relacionados otros dos grupos $\bar{G}$ y G' de transformaciones de los espacios Θ y D en sí. El empleo simultáneo de las tres transformaciones g, $\bar{g}$ y g' deja inalterable (invariante) el problema de decisión. Por eso es natural elegir tales reglas de decisión que no varíen al pasar de un problema de decisión equivalente a otro. En los §§ 2.18, 2.19 y 3.7 ya hemos analizado muy detalladamente la naturaleza de tal enfoque.

Definición 3. La función de decisión $\delta(X)$ del problema invariante de decisión se llama *invariante* si

$$\delta(gX) = g'\delta(X).$$

La *regla invariante randomizada* $\pi(X)$ se define como cualquier distribución concentrada en las reglas invariantes de decisión.

Ejemplos de utilización del principio de invariación se ofrecen en los §§ 2.18, 2.19 y 3.7 ya mencionados, donde hemos examinado las estimaciones equivariantes y los criterios invariantes. Es preciso señalar cierta peculiaridad de estos dos casos particulares desde el punto de vista del enfoque general.

En el *problema de estimación*, el grupo de transformación G' no se ha introducido en absoluto. En este caso, los conjuntos D y Θ coinciden, y desde el principio se suponía que $g'\delta = \bar{g}\delta$. Por eso hemos definido las estimaciones equivariantes con ayuda de la igualdad $\theta^*(gX) = \bar{g}\theta^*(X)$.

En la *teoría de verificación de hipótesis* se suponía que la transformación g' era igual a la transformación idéntica $g' = e$, por lo tanto, el criterio invariante π podía ser definido por la relación $\pi(gX) = \pi(X)$.

En este caso, para la invariación del problema de verificación de dos hipótesis $\{\theta \in \Theta_1\}$ y $\{\theta \in \Theta_2\}$ también es necesario suponer (véase (4)) que $g\Theta_i = \Theta_i$.

Precisamente debido a la existencia de cierta diferencia en estos dos enfoques se explica, en cierta medida, la utilización de dos términos diferentes: "equivariación" (para las estimaciones) e "invariación" (para la verificación de hipótesis) para designar las reglas de decisión invariantes. Adicionalmente a los ejemplos de problemas invariantes de decisión, examinados en los capítulos 2 y 3, citaremos uno más.

Ejemplo 1. Supongamos que $X \in \Phi_{\alpha,\sigma^2}$. Aquí Θ es el semiplano $\{\theta = (\alpha, \sigma): \sigma \geqslant 0\}$. Sea D la recta real R, y sea $w(\delta, \theta) = (\delta - \alpha)^2/\sigma^2$.

Examinemos el grupo G de transformaciones $g_{a,b}X = a + bX = (a + bx_1, \dots, a + bx_n)$, donde $b \neq 0$. La variable aleatoria $g_{a,b}X$ en $\mathscr{X}^n$ puede, evidentemente, considerarse como una muestra de $\Phi_{a+b\alpha,\, b^2\sigma^2}$. Por consiguiente, la familia Φ_{α,σ^2} es invariante respecto a G, si se pone $\bar{g}_{a,b}\theta = (a + b\alpha,\ |b|\sigma)$. La función de pérdidas será invariante si ponemos $g'_{a,b}\delta = a + b\delta$, puesto que

$$w(g'_{a,b}\delta,\ \bar{g}_{a,b}\theta) = \frac{(a + b\delta - a - b\alpha)^2}{b^2\sigma^2} = w(\delta, \theta).$$

Ahora bien, tenemos un problema invariante de decisión respecto a G. Las funciones invariantes de decisión $\delta(X)$: $\mathscr{X}^n \to R$ deben poseer la propiedad

$$\delta(a + bX) = \delta(g_{a,b}X) = g'_{a,b}\delta(X) = a + b\delta(X). \tag{5}$$

Seguidamente, no es difícil establecer que el problema de decisión sometido a examen también es invariante respecto al grupo F de todas las permutaciones f de las coordenadas del vector X; en este caso, $\bar{f}$ y f' serán dos transformaciones idénticas. Por eso, si exigimos, que la función $\delta(X)$ también sea una decisión invariante respecto a F, entonces también debe cumplirse

$$\delta(fX) = \delta(X). \tag{6}$$

Nótese que la clase de funciones que satisfacen (5) y (6) aún es bastante amplia: en ella entran, por ejemplo, todas las formas lineales

$$\delta(X) = \sum_{k=1}^{n} a_k x_{(k)}, \quad \sum_{k=1}^{n} a_k = 1,$$

donde $x_{(1)}, \dots, x_{(n)}$ es la serie variacional de la muestra X. Si utilizamos el principio de no desplazamiento, obtendremos una condición más para los coeficientes a_k:

$$\sum_{k=1}^{n} a_k \mathbf{M}_\theta(x_{(k)} - \alpha) = 0. \ \triangleleft$$

Al construir las decisiones invariantes óptimas en la teoría de estimación y en la teoría de verificación de las hipótesis estadísticas, desempeñan un papel muy importante los conceptos que, en cierto sentido, se asemejan uno a otro: el concepto de órbita en la teoría de estimaciones, y el concepto de invariante en la teoría de verificación de hipótesis. Recordemos que por órbita en el espacio Θ se entiende el conjunto $\{\bar{g}\theta_0, \bar{g} \in \bar{G}\}$, donde θ_0 es cierto punto de Θ. Con otras palabras, θ_1 y θ_2 pertenecen a una misma órbita, si existe $\bar{g} = \bar{C}$ tal, que $\theta_1 = \bar{g}\theta_2$.

Análogamente se pueden definir las órbitas en $\mathscr{X}^n$. Entonces son invariantes, por definición, las estadísticas constantes en las órbitas en $\mathscr{X}^n$.

El concepto de órbita también conserva su importancia en el caso general.

Lema 2. *La función de riesgo del problema invariante de decisión para una regla invariante de decisión, es constante en la órbita:*

$$W(\delta(\cdot),\ \theta) = W(\delta(\cdot),\ \bar{g}\theta)$$

para todos $\theta \in \Theta$, $\bar{g} \in \bar{G}$.

Demostración. En virtud de la invariación respectiva de la función de pérdidas, de la regla de decisión y de la familia $\mathbf{P}_\theta$ (véanse (3) y (4)), tenemos

$$\begin{aligned} W(\delta(\cdot),\ \theta) &= \mathbf{M}_\theta w(\delta(X),\ \theta) = \mathbf{M}_\theta w(g'\,\delta(X),\ \bar{g}\theta) = \\ &= \mathbf{M}_\theta w(\delta(gX),\ \bar{g}\theta) = \mathbf{M}_{\bar{g}\theta} w(\delta(X),\ \bar{g}\theta) = W(\delta(\cdot),\ \bar{g}\theta). \ \lhd \end{aligned}$$

La constancia en la órbita de riesgo para las reglas de decisión invariantes randomizadas se deduce de su definición y del lema 2.

De este último resulta que en el caso de que todo el espacio Θ sea una órbita (es decir, $\Theta = \{\bar{g}\theta_0, \bar{g} \in \bar{G}\}$ para cualquier θ_0; esto tiene lugar, por ejemplo, para las transformaciones de desplazamiento), la regla invariante de decisión será una regla igualadora. Por eso, del lema 2 y de los teoremas 2.3, 2.5 obtenemos directamente la siguiente afirmación que establece una relación importante entre la invariación y el carácter minimax.

Teorema 2. *Supongamos que el espacio Θ es una órbita y que existe una distribución a priori $\mathbf{Q}$ para la cual la estrategia bayesiana $\pi_Q(X)$ es invariante. Entonces $\pi_Q(X)$ será una estrategia minimax.*

Del teorema 3.3 se desprende que tiene lugar la siguiente generalización del teorema 2.

Teorema 2A. *Supongamos que existe una distribución a priori $\mathbf{Q}$ concentrada en una de las órbitas, tal, que la estrategia Θ_0 bayesiana $\pi_Q(X)$ es invariante.*

Entonces, si para todos θ,

$$W(\pi_Q(\cdot),\ \theta) \leqslant W(\pi_Q(\cdot),\ \theta_0), \quad \theta_0 \in \Theta_0,$$

entonces $\pi_Q(X)$ es minimax.

Este criterio fue utilizado en el § 3.9.

§ 6. Estimaciones asintóticamente óptimas para una función de pérdidas arbitraria

Muchos de los resultados de las estimaciones asintóticamente óptimas (capítulo 2) y de los criterios asintóticamente óptimos (capítulo 3) admiten generalizaciones en la función de pérdidas, de forma muy general.

En este párrafo investigaremos los problemas de la teoría de estimación y supondremos que $w(\delta, \theta) = w(\delta - \theta)$.

Hagamos primeramente una observación general. En el capítulo 2 hemos visto que en el caso general ($X \in \mathbf{P}_\theta$, $\mathbf{P}_\theta$ satisface las condiciones (RR); véanse los §§ 2.24 y 2.28), todas las estimaciones racionales $\theta^* = \delta(X)$ del parámetro θ están "concentradas" en el entorno $1/\sqrt{n}$ del punto θ. Así, por ejemplo, para las estimaciones asintóticamente normales, $(\theta^* - \theta)\sqrt{n} \Subset \Phi_{0,\sigma^2(\theta)}$. De aquí se deduce que, para amplias suposiciones respecto a la función $w(t)$, el comportamiento asintótico del riesgo $\mathbf{M}_\theta w(\theta^* - \theta)$ será determinado por las propiedades de la función $w(t)$ en el entorno del punto $t = 0$. Si $w(t)$ es dos veces continuamente derivable en el cero, $w'' > 0$, entonces, para $t \to 0$,

$$w(t) = \frac{w''(0)}{2} t^2 + o(t^2). \tag{1}$$

Esto significa que en la región de valores de t (del orden de $1/\sqrt{n}$) que nos interesa, la función $w(t)$ se comportará igual que la función cuadrática de pérdidas $w_0(t) = ct^2$, cuando $c = w''(0)/2$, para la que han sido establecidos los resultados del capítulo 2. Si, además, $w(t) < e^{\alpha |t|^2}$, siendo bastante pequeño $\alpha > 0$ (véase el teorema 2.28.6), todos estos resultados mantendrán su validez, ya que su traslado al caso de la función $w(t)$ de forma (1), es cuestión de una técnica no complicada, completamente al alcance del lector.

En este párrafo examinaremos una generalización mucho más sustancial. Supondremos que la función de pérdidas $w(\delta, \theta)$ dependa de n y que la misma es representable en la forma

$$w(\delta, \theta) = w_n(\delta - \theta) = w(\sqrt{n}(\delta - \theta)), \tag{2}$$

donde la función $w(t) \geqslant 0$ está definida en todo el espacio R^k. Es evidente que en este caso serán esenciales los valores de $w(t)$ en toda la región de los valores de t.

Admitiremos que la función w en (2) satisface las condiciones siguientes:

1) $w(t) \leqslant e^{c|t|}$ *para cierto* $c > 0$.

Tal forma de condición 1) simplifica algo los cálculos. En efecto, todos los resultados conservarán su validez si exigimos que $w(t) \leqslant c_1 e^{\alpha |t|^2}$ cuando $\alpha > 0$ es bastante pequeño.

Posteriormente desempeñará un papel muy importante la función

$$V_{\sigma^2}(s) = \int w(s-u)e^{-\frac{1}{2}u\sigma^2u^T}du,$$

donde σ^2 es cierta matriz de segundos momentos, definida positivamente. La función $V_{\sigma^2}(s)$ puede interpretarse como

$$V_{\sigma^2}(s) = \frac{(2\pi)^{k/2}}{\sqrt{|\sigma^2|}}\mathbf{M}w(s-\xi), \quad \xi \in \Phi_{0,\sigma^{-2}}.$$

En vista de que

$$V_{\sigma^2}(s) = \int w(v)e^{-\frac{1}{2}(s-v)\sigma^2(s-v)^T}dv,$$

esta función será la función analítica de las variables s y σ^2.

También necesitaremos las condiciones:

2) *La función $V_{\sigma^2}(s)$ alcanza su valor mínimo respecto a s en un solo punto que designaremos por b_w.*

3) $b_w = 0$.

4) *La función $w(t)$ es continua.*

La condición 2) se cumplirá a ciencia cierta si $w(s) \neq$ const es una función convexa hacia abajo. En este caso $V_{\sigma^2}(s)$ será, evidentemente, también convexa y no contendrá partes "lineales" (o sea, la matriz de segundas derivadas será por doquier definida positivamente).

La condición 3) será cumplida si

$$V'_{\sigma^2}(0) = -\int uw(u)e^{-\frac{1}{2}u\sigma^2u^T}du = 0,$$

lo cual siempre tendrá lugar para las funciones simétricas $w(u) = w(-u)$.

El valor de b_w podría llamarse *desplazamiento* de la función de pérdida w. El mismo satisface la ecuación $V'_{\sigma^2}(b_w) = 0$. La condición 3) acerca de que $b_w = 0$ no es esencial y sólo simplifica la exposición, que el lector también puede extender fácilmente al caso de $b_w \neq 0$. Las modificaciones que en este caso tendrán lugar en los enunciados de los teoremas, serán ilustradas en la observación 2 correspondiente al teorema 1.

Recordemos ahora en qué se transformarán las definiciones de las estrategias óptimas expuestas en los §§ 2 y 3. La estimación θ^*_Q será *bayesiana* respecto a la distribución a priori $\mathbf{Q}$ con densidad q respecto a la medida de Lebesgue (y a la función de pérdidas w_n) si

$$\int W(\theta^*_Q, t)q(t)dt = \min_{\theta^*} \int W(\theta^*, t)q(t)dt, \tag{3}$$

donde $W(\theta^*, t) = \mathbf{M}_t w_n(\theta^* - t)$. Aquí la integral del segundo miembro (3) puede escribirse en forma de la esperanza matemática incondicional $\mathbf{M}w_n(\theta^* - \theta)$, donde la promediación se toma respecto a la distribución con densidad $f_t(x)q(t)$.

La estimación $\bar{\theta}^*$ será *minimax* si para cualquier otra estimación θ^*,

$$\sup_t W(\bar{\theta}^*, t) \leqslant \sup_t W(\theta^*, t).$$

Lo dicho hace naturales las siguientes definiciones que son completamente análogas a las dadas en el § 2.11.

Definición 1. Llamaremos *asintóticamente bayesiana* la estimación θ^* si

$$\varlimsup_{n\to\infty} \sup\,[\mathbf{M}w_n(\theta^* - \theta) - \mathbf{M}w_n(\theta_Q^* - \theta)] \leqslant 0, \tag{4}$$

donde θ_Q^* es la estimación bayesiana.

Definición 2. Llamaremos *asintóticamente minimax* la estimación θ_1^*, si para cualquier otra estimación θ^*,

$$\varlimsup_{n\to\infty} \sup\left[\sup_{t\in\Theta_0} W(\theta_1^*, t) - \sup_{t\in\Theta_0} W(\theta^*, t)\right] \leqslant 0, \tag{5}$$

donde Θ_0 es cualquier subconjunto cerrado que se encuentra dentro de Θ.

Al estudiar las estimaciones asintóticamente óptimas en este párrafo, sólo utilizaremos los conceptos introducidos en las definiciones 1 y 2. Esto constituye cierta diferencia del capítulo 2, donde también estaban presentes las estimaciones asintóticamente eficientes. Aquí su ausencia se explica por el hecho de que para las funciones arbitrarias de pérdidas w no disponemos de desigualdades del tipo de Rao — Cramer para $\inf\limits_{\theta^*\in K_0} W(\theta^*, \theta)$ (K_0 es la clase de estimaciones no desplazadas), con ayuda de la cual era posible, valiéndose del valor de $W(\theta^*, \theta)$, juzgar acerca de la calidad de θ^* y destacar, en particular, las estimaciones eficientes (y asintóticamente eficientes), o sea, las estimaciones uniformemente mejores en la clase K_0.

Las afirmaciones siguientes establecen que la estimación de verosimilitud máxima es, al igual que en las condiciones del capítulo 2, asintóticamente bayesiana y asintóticamente minimax. Además, obtendremos la frontera inferior *asintótica* para la función de riesgo al ser arbitraria la función de pérdidas w (la desigualdad de Rao — Cramer proporciona la frontera inferior *exacta*). En los tres teoremas ulteriores supondremos que se cumple la condición (RR).

Teorema 1. *Supongamos que $X \in \mathbf{P}_\theta$, θ^* es la e.v.m., y que θ_Q^* es una estimación bayesiana correspondiente a la función de pérdidas w* (véase (2)) *que satisface las condiciones* 1) — 3), *así como a la distribución a priori* $\mathbf{Q}$ *con una densidad q limitada respecto a la medida de Lebesgue. Entonces*

$$|\theta_Q^* - \hat{\theta}^*|\sqrt{n} \xrightarrow[P_\theta]{} 0, \tag{6}$$

$$(\theta_Q^* - \theta)\sqrt{n} \in \Phi_{0,I^{-1}(\theta)} \tag{7}$$

es uniforme respecto a $\theta \in \Theta_0$; Θ_0 cualquier subconjunto cerrado, situado dentro de Θ, en el que $q(\theta) > q_0 > 0$ es continua.

Si, además, la función w satisface la condición (4), *entonces*

$$\mathbf{M}w_n(\theta_Q^* - \theta) = \mathbf{M}w(\sqrt{n}(\theta_Q^* - \theta)) \to \mathbf{M}w(\eta_\theta) = \mathbf{M}\,\frac{\sqrt{I(0)}}{(2\pi)^{K/2}}\,V_{I(\theta)}(0), \qquad (8)$$

donde $\eta_\theta \in \Phi_{0,I^{-1}(\theta)}$, $\theta \in \mathbf{Q}$; $\mathbf{M}$, *como antes, designa la esperanza matemática incondicional cuya densidad constituye* $f_t(x)\,q(t)$ $(X \in \mathbf{P}_\theta,\ \theta \in \mathbf{Q})$.

Observación 1. A la par con la convergencia (6) también se puede establecer una convergencia casi segura respecto a $\mathbf{P}_\theta$.

Observación 2. Si w es tal que el desplazamiento $b_w \neq 0$, la afirmación del teorema 1 quedará válida por completo, siempre que θ_Q^* en (6), (7) y (8) se sustituya por $\theta_Q^* - b_w/\sqrt{n}$. Ahora bien, b_w tiene sentido de *desplazamiento asintótico de la magnitud* $(\theta_Q^* - \theta)\sqrt{n}$.

Teorema 2. *Supongamos que la función w satisface las condiciones* 1) — 4). *Entonces, para cualquier estimación* θ^*,

$$\lim_{n\to\infty}\ \inf\ \sup_{t\in\Theta_0} \mathbf{M}_t w_n(\theta^* - t) \geqslant \sup_{t\in\Theta_0} \mathbf{M}w(\eta_t), \qquad (9)$$

$$\eta_t \in \Phi_{0,I^{-1}(t)}.$$

Cualquier estimación θ^* *para la cual*

$$\mathbf{M}_t w_n(\theta^* - t) \to \mathbf{M}w(\eta_t) \qquad (10)$$

uniformemente respecto a t, es asintóticamente minimax.

Teorema 3. *Supongamos que* $X \in \mathbf{P}_\Theta$ *y que la función w satisface las condiciones* 1) — 4). *Entonces, la estimación de verosimilitud máxima* $\hat{\theta}^*$ *es asintóticamente minimax y asintóticamente bayesiana para cualquier distribución a priori* $\mathbf{Q}$ *cuya densidad q es continuamente positiva en el punto* θ.

Todas estas afirmaciones son absolutamente análogas a las afirmaciones correspondientes del capítulo 2, ya que las mismas contribuyen a la verosimilitud de la suposición de que también para la función de pérdidas arbitraria w que satisface las condiciones 1) — 4), la e.v.m. es la mejor estimación asintóticamente uniforme en la clase de estimaciones asintóticamente no desplazadas (compárese con los §§ 2.25 y 2.28).

Demostración del teorema 1. En virtud del principio bayesiano, la estimación bayesiana se define como el valor θ_Q^* que posee la propiedad

$$\int w_n(\theta_Q^* - t)q(t/X)dt = \min_{u\in\Theta} \int w_n(u - t)q(t/X)dt =$$

$$= \min_{u\in\Theta} \int w(\sqrt{n}(u - \theta) - \sqrt{n}(t - \theta))\,\frac{q(t)f_t(X)}{\int q(v)f_v(X)dv}\,dt.$$

Esto significa que en calidad de $(\theta_Q^* - \theta)\sqrt{n} \equiv u_Q^*$ se puede tomar cualquier valor s con el cual se alcanza mín $U(s)$,

$$U(s) = \int w(s - v)q\left(\theta + \frac{v}{\sqrt{n}}\right) Z\left(\frac{v}{\sqrt{n}}\right) dv, \tag{11}$$

donde, como antes, $Z(t) = \dfrac{f_{\theta+t}(X)}{f_\theta(X)}$.

Necesitaremos las afirmaciones acerca del comportamiento asintótico de $U(s)$. En los §§ 2.28 y 2.29 hemos establecido (teorema 2.28.5) que, al cumplirse las condiciones (RR),

$$U(u^*) = e^{Y(u^*)}q(\hat{\theta}^*)(V_{I(\theta)}(0) + \varepsilon_n(X, \theta)), \tag{12}$$

donde $\varepsilon_n(X, \theta) \underset{P}{\to} 0$ uniformemente respecto a θ (aquí hemos sustituido $\dfrac{(2\pi)^{k/2}}{\sqrt{I(\theta)}} \mathbf{M}w(\xi)$ por $V_{I(\theta)}$, y $q(\theta)$ por $q(\hat{\theta}^*)$).

Nótese ahora que

$$\mathbf{P}(\sqrt{n}\,|\theta_Q^* - \hat{\theta}^*| \geqslant \varepsilon) = \mathbf{P}(|u_Q^* - u^*| \geqslant \varepsilon) \leqslant \\ \leqslant \mathbf{P}\left(\min_{|s-u^*|\geqslant\varepsilon} U(s) \leqslant U(u^*)\right). \tag{13}$$

En vista de que tenemos la representación asintótica para $U(u^*)$, aquí debemos estimar el valor de $U(s)$. De los teoremas 2.28.4 y 2.29.3 se deduce que para la sucesión arbitraria $\delta_n \to 0$, cuando $|v| < \delta_n\sqrt{n}$,

$$\ln Z\left(\frac{v}{\sqrt{n}}\right) = Y(u^*) - \frac{1}{2}(v - u^*)I(\theta)(v - u^*)^T(1 + \varepsilon_n(X, \theta, u)),$$

$|\varepsilon_n(X, \theta, u)| \leqslant \varepsilon_n^{(1)}(X, \theta) \underset{P_\theta}{\to} 0$ uniformemente respecto θ. Pero

$$U(s) \geqslant U_n(s) \equiv \int_{|v-u^*|\leqslant\delta_n\sqrt{n}} w(s - v)q\left(\theta + \frac{v}{\sqrt{n}}\right) Z\left(\frac{v}{\sqrt{n}}\right) dv.$$

Examinemos el conjunto

$$A_n = \left\{\varepsilon_n^{(1)}(X, \theta) < \varrho, \inf_{\substack{|v-u^*|\leqslant\delta_n\sqrt{n} \\ \varrho>0,}} q\left(\theta + \frac{v}{\sqrt{n}}\right) > q(\hat{\theta}^*)(1 - \varrho)\right\},$$

que posee, evidentemente, la propiedad

$$\mathbf{P}_\theta(A_n) \to 1. \tag{14}$$

En este conjunto, uniformemente respecto a θ,

$$U_n(s) \geqslant (1 - \varrho)q(\hat{\theta}^*)e^{Y(u^*)} \times \\ \times \int_{|v-u^*|\leqslant\delta_n\sqrt{n}} w(s - v)\exp\left\{-\frac{1}{2}(v - u^*)I(\theta)(v - u^*)^T(1 + \varrho)\right\} dv = \\ = (1 - \varrho)q(\hat{\theta}^*)e^{Y(u^*)}[V_{I(\theta)(1+\varrho)}(s - u^*) - r_n(s)], \tag{15}$$

donde, según la condición 1),

$$r_n(s) = \int\limits_{|v-u^*|\leqslant\delta_n\sqrt{n}} w(s-v)\exp\left\{-\frac{1}{2}(v-u^*)I(\theta)(v-u^*)^T\times\right.$$
$$\left.\times(1+\varrho)\right\}dv \leqslant e^{c\sqrt{nd}}\frac{(2\pi)^{k/2}}{\sqrt{|I(\theta)|c}}\mathbf{P}(|\eta|>\delta_n\sqrt{n}),$$
$$\eta \in \Phi_{0,I(\theta)(1+\varrho)},$$

donde d es el diámetro de la región Θ. Al igual que en el lema 2.23.1, es fácil convencerse de que

$$\mathbf{P}(|\eta|>\delta_n\sqrt{n}) \leqslant e^{-\alpha n\delta_n^2},\ \alpha>0.$$

Eligiendo $\delta_n = n^{-1/9}$, obtenemos que, para todos los valores de s y con valores de n bastante grandes,

$$r_n(s) \leqslant e^{-n^{3/4}}. \tag{16}$$

Ahora utilicemos las condiciones 2) y 3) en virtud de las cuales

$$\min_{|s-u^*|\geqslant\varepsilon} V_{I(\theta)}(s-u^*) \geqslant V_{I(\theta)}(0)+4\tau,\ \tau=\tau(\varepsilon)>0.$$

En virtud de las propiedades analíticas de $V_{\sigma^2}(s)$ obtendremos que, para valores de ϱ bastante pequeños,

$$\min_{|s-u^*|\geqslant\varepsilon} V_{I(\theta)(1+\varrho)}(s-u^*) \geqslant V_{I(\theta)}(0)+3\tau,$$

y en virtud de (15) y (16), para $X\in A_n$ y para valores de n bastante grandes,

$$\min_{|s-u^*|\geqslant\varepsilon} U_n(s) \geqslant (1-\varrho)q(\hat{\theta}^*)e^{Y(u^*)}[V_{I(\theta)}(0)+2\tau].$$

Utilizando (12) y (13), definitivamente obtenemos

$$\mathbf{P}_\theta(\sqrt{n}|\theta_Q^*-\hat{\theta}^*|\geqslant\varepsilon) \leqslant \mathbf{P}_\theta(\min_{|s-u^*|\geqslant\varepsilon} U_n(s)\leqslant U(u^*)) \leqslant$$
$$\leqslant \mathbf{P}_\theta(X\notin A_n)+\mathbf{P}_\theta((1-\varrho)[V_{I(\theta)}(0)+2\tau]\leqslant V_{I(\theta)}(0)+\varepsilon_n(X,\Theta)).$$

Eligiendo adicionalmente ϱ, de tal modo que su valor sea tan pequeño que contribuya al cumplimiento de $(1-\varrho)2\tau - V_{I(\theta)}(0)\geqslant\tau$, obtendremos

$$\mathbf{P}(\sqrt{n}|\theta_Q^*-\hat{\theta}^*|\geqslant\varepsilon) \leqslant \mathbf{P}_\theta(X\notin A_n)+\mathbf{P}_\theta(\varepsilon_n(X,\theta)>\tau)\to 0$$

cuando $n\to\infty$. En virtud de (12) y (14), la afirmación (6) queda demostrada.

De (6) y de los teoremas del § 2.29 se desprende (7). Demostremos ahora la relación (8). En virtud de (7) y de la propiedad (4),

$$w(\sqrt{n}(\theta_Q^*-\theta)) \Rightarrow w(\eta_\theta),\quad \eta_\theta\in\Phi_{0,I^{-1}(\theta)}.$$

Según el lema de Fatou,

$$\lim_{n\to\infty} \inf \mathbf{M}_t w(\sqrt{n}(\theta_Q^* - t)) \geqslant \mathbf{M} w(\eta_t),$$

$$\lim_{n\to\infty} \inf \mathbf{M} w(\sqrt{n}(\theta_Q^* - \theta)) \geqslant \int q(t)\mathbf{M} w(\eta_t)dt = \mathbf{M} w(\eta_t) = \mathbf{M} w(\eta_\theta).$$

Por otro lado, según la definición de θ_Q^*,

$$\mathbf{M} w(\sqrt{n}(\theta_Q^* - (\theta)) \leqslant \mathbf{M} w(\sqrt{n}(\hat{\theta}^* - \theta)) \to \mathbf{M} w(\eta_\theta).$$

La última relación se deduce de la convergencia uniforme $\mathbf{M}_t w(\sqrt{n}(\theta^* - t)) \to \mathbf{M} w(\eta_t)$ demostrada en el § 2.29. El teorema queda demostrado.

Demostración del teorema 2. Tomemos la distribución $\mathbf{Q}$ concentrada en Θ_0, con una densidad limitada $q(t) > 0$ para $t \in \Theta_0$, y sea θ_Q^* la estimación bayesiana correspondiente a $\mathbf{Q}$. Entonces, para cualquier estimación θ^*,

$$\sup_{t\in\Theta_0} \mathbf{M}_\theta w_n(\theta^* - t) \geqslant \int_{\Theta_0} \mathbf{M}_t w_n(\theta^* - t)q(t)dt \geqslant$$

$$\geqslant \int_{\Theta_0} \mathbf{M}_t w_n(\theta_Q^* - t)q(t)dt = \mathbf{M} w_n(\theta_Q^* - \theta).$$

Según el lema de Fatou, en virtud de (8),

$$\lim_{n\to\infty} \inf \sup_{t\in\Theta_0} \mathbf{M}_t w_n(\theta^* - t) \geqslant \lim_{n\to\infty} \inf \mathbf{M} w_n(\theta_Q^* - \theta) \geqslant \mathbf{M} w(\eta_\theta) =$$

$$= \int_{\Theta_0} \mathbf{M} w(\eta_t)q(t)dt.$$

Como la función $\mathbf{M} w(\eta_t) = \dfrac{\sqrt{I(t)}}{(2\pi)^{k/2}} V_{I(t)}(0)$ es continua respecto a t, entonces, eligiendo $q(t)$, podemos conseguir que la integral

$$\int_{\Theta_0} \sqrt{I(t)}\, V_{I(t)}(0)q(t)dt$$

se asemeje tanto como se quiera a $\sup_{t\in\Theta_0} \sqrt{I(t)}\, V_{I(t)}(0) = \sup_{t\in\Theta_0} \mathbf{M} w(\eta_t)$. Esto demuestra (9).

Ahora supongamos que la estimación θ_1^* posee la propiedad (10), y que θ^* es cualquier otra estimación. Entonces, en virtud de (9) y de la convergencia uniforme (10),

$$\lim_{n\to\infty} \sup \left[\sup_{t\in\Theta_0} \mathbf{M}_t w_n(\theta_1^* - t) - \sup_{t\in\Theta_0} \mathbf{M}_t w_n(\theta^* - t)\right] \leqslant$$

$$\leqslant \sup_{t\in\Theta_0} \lim_{n\to\infty} \mathbf{M}_t w_n(\theta_1^* - t) - \sup_{t\in\Theta_0} \mathbf{M}_w(\eta_t) = 0.$$

La desigualdad (5) de definición del carácter asintóticamente minimax, y junto con ella el teorema 2, quedan demostrados.

Demostración del teorema 3. El carácter asintóticamente minimax de $\hat{\theta}^*$ se desprende del hecho de que para la e.v.m. $\hat{\theta}^*$, según el teorema 2.29.4, es válida (10).

El carácter asintóticamente bayesiano de $\hat{\theta}^*$ se deduce del hecho de que para $\theta^* = \hat{\theta}^*$ se cumple (4), ya que para $\hat{\theta}^*$ tiene lugar la convergencia uniforme (10) y, por lo tanto,

$$\lim_{n\to\infty} \mathbf{M}w_n(\hat{\theta}^* - \theta) = \lim_{n\to\infty} \int \mathbf{M}_t w_n(\hat{\theta}^* - \theta)q(t)dt =$$
$$= \mathbf{M}w(\eta_\theta) = \lim_{n\to\infty} \mathbf{M}w_n(\theta_Q^* - \theta).$$

La última igualdad resulta de (8). El teorema queda demostrado.

La afirmación del teorema 1 puede ser reforzada si se exige adicionalmente que la función $w(t)$ aumente con bastante rapidez. Para esto, designemos $w_N = \min\limits_{|t|>N} w(t)$ y $W_M = \max\limits_{|t|\leq M} w(t)$ y examinemos la condición

5) *Existe $\gamma < 1$ tal, que $w_N > 2W_{\gamma N}$ para todos los valores de N bastante grandes.*

Si cuando $|t| \to \infty$, $w(t)$ crece como función potencial o exponencial, entonces se cumple la condición 5).

Teorema 4. *Si se cumplen las condiciones* 1) *y* 5) *cuando $q(t) > q_0 > 0$ en el conjunto cerrado Θ_0, y cuando $q(t) \leq q_m < \infty$, entonces, para ciertos valores de $c < \infty$ y de $\alpha > 0$ que no dependen de t,*

$$\mathbf{P}t(\sqrt{n}(\theta_Q^* - t) > N) \leq ce^{-\alpha N^2}, \quad t \in \Theta_0.$$

De aquí y del teorema 1 se deduce que para cualquier función continua $v(t)$ tal, que $|v(t)| \leq e^{-\alpha N^2/2}$, es válida

$$\mathbf{M}_t v(\sqrt{n}(\theta_Q^* - t)) \to \mathbf{M}v(\eta_t), \quad t \in \Theta_0.$$

Designemos

$$u(r) = \int\limits_{|v|\geq r} w(-v)q\left(\theta + \frac{v}{\sqrt{n}}\right)Z\left(\frac{v}{\sqrt{n}}\right)dv$$

(ésta es la parte de la integral $U(0)$ que se encuentra en la región $|v| \geq r$. Para demostrar el teorema 4 necesitaremos el

Lema 1. *Si $w(t)$ satisface la condición* 1), *y $q_m = \max\limits_u q(u) < \infty$, entonces, para ciertos $\beta > 0$ y $a < \infty$ que no dependen de θ, así como para todos $0 < \delta < 1$,*

$$\mathbf{P}_\theta(u(r) > \delta) \leq \frac{a}{\delta}e^{-\beta r^2}.$$

Esta desigualdad quedará válida para $w(t) \equiv 1$.

Demostración. Tenemos

$$\mathbf{P}_\theta(u(r) > \delta) \leq \mathbf{P}_\theta\left(\sup_{|v|\geq r} Z\left(\frac{v}{\sqrt{n}}\right) > 1 + \mathbf{P}_\theta(u(r) > \delta, \sup_{|v|\geq r} Z\left(\frac{v}{\sqrt{n}}\right) \leq 1\right).$$

La estimación del primer sumando se da en el teorema 2.23.2, en virtud de la cual este sumando no pasa de $c_1e^{-r^2\beta}$, $\beta > 0$. El segundo sumando no supera

$$\mathbf{P}_\theta \left(\int\limits_{|v|\geqslant r} w(-v)q\left(\theta + \frac{v}{\sqrt{n}}\right) Z^{1/2}\left(\frac{v}{\sqrt{n}}\right) dv > \delta \right). \tag{17}$$

Como, en virtud del teorema 2.23.1,

$$\mathbf{M}_\theta Z^{1/2}\left(\frac{v}{\sqrt{n}}\right) \leqslant e^{-2|v|^2\beta},\ \beta > 0,$$

la esperanza matemática de la integral en (17) no superará (véase el lema 2.23.1.)

$$q_m \int\limits_{|v|\geqslant r} e^{c|v|}e^{-2|v|^2\beta}dv \leqslant c_2e^{-r^2\beta}.$$

Por eso, en virtud de la desigualdad de Chébishev, la probabilidad (17) no supera $c_2e^{-r^2\beta}/\delta$. El lema queda demostrado.

Designemos por u_1r el valor de la integral $u(r)$ cuando $w(t) \equiv 1$:

$$u_1(r) = \int\limits_{|v|\geqslant r} q\left(0 + \frac{v}{\sqrt{n}}\right) Z\left(\frac{v}{\sqrt{n}}\right) dv.$$

Lema 2. *Si $q(\theta) > 0$ en el conjunto cerrado Θ_0, entonces, con cierto $b < \infty$ que no depende de θ, para cualquer $\varepsilon > 0$ y para todos los valores de n bastante grandes,*

$$\mathbf{P}_\theta(u_1(0) < \varepsilon) \leqslant b\varepsilon^2,\quad \theta \in \Theta_0.$$

Demostración. Para todos los valores de n bastante grandes tenemos

$$u_1(0) \geqslant \int\limits_{|v|\leqslant 1} q\left(\theta + \frac{v}{\sqrt{n}}\right) Z\left(\frac{v}{\sqrt{n}}\right) dv \geqslant$$

$$\geqslant q_0 \int\limits_{|v|\leqslant 1} \exp\left\{L\left(X,\ \theta + \frac{v}{\sqrt{n}}\right) - L(X,\ \theta)\right\} dv =$$

$$= q_0 \int\limits_{|v|\leqslant 1} \exp\left\{(v,\ \zeta_n) + \frac{1}{2}\, v\gamma_n v^T\right\} dv,$$

donde

$$q_0 = \min_{t\in\Theta_0} q(t) > 0,\ \zeta_n = \frac{1}{\sqrt{n}} L'(X,\ t),\ \gamma_n = \frac{1}{n}\,\| L_{ij}''(X,\ \tilde{\theta})\|,$$

$\tilde{\theta} = \theta + \varrho v n^{-1/2}$, $|\varrho| \leqslant 1$. (Aquí L' es el vector de las derivadas de la función logarítmica de verosimilitud; L_{ij}'', las derivadas parciales de segundo orden.) En vista de que $|v,\ \zeta_n| \leqslant |v||\zeta_n|$ y como, en virtud de las condiciones (RR),

$$|v\gamma_n v^T| \leqslant \frac{1}{n}\sum_{i=1}^{n} l(x_i) \sum_{i,j=1}^{k} |v_i v_j| \leqslant \frac{k|v|^2}{n} L_n,$$

donde $L_n = \sum\limits_{i=1}^{n} l(x_i)$, entonces, en el conjunto $A = \{|\zeta_n| \leqslant 1/\varepsilon,\ L_n \leqslant n/\varepsilon^2 k\}$ es válida

$$u_1(0) \geqslant q_0 \int\limits_{|v|\leqslant 1} \exp\left\{-\frac{|v|}{\varepsilon} - \frac{|v|^2}{2\varepsilon^2}\right\} dv \geqslant q_0\varepsilon \int\limits_{|s|\leqslant \varepsilon^{-1}} \exp\left\{-|s| - \frac{|s|^2}{2}\right\} ds \geqslant c_1\varepsilon.$$

Esto quiere decir que tiene lugar el encaje $\{u_1(0) < c_1\varepsilon\} \subset \bar{A}$. Como

$$\mathbf{P}_\theta(\bar{A}) \leqslant \mathbf{P}_\theta(|\zeta_n| \geqslant \varepsilon^{-1}) + \mathbf{P}_\theta\left(L_n > \frac{n}{\varepsilon^2 K}\right) \leqslant \varepsilon^2 \mathbf{M}_\theta |\zeta_n|^2 + \frac{\varepsilon^2 k}{n} \mathbf{M}_\theta L_n,$$

$$\mathbf{M}_\theta |\zeta_n|^2 = \sum_{i=1}^{k} I_{ii}(\theta), \ \mathbf{M}_\theta L_n = n\mathbf{M}_\theta l(x_1),$$

entonces

$$\mathbf{P}_\theta(\bar{A}) \leqslant c_2\varepsilon^2.$$

El lema queda demostrado.

Demostración del teorema 4. Designemos por M_n el conjunto de puntos s en los cuales se alcanza mín $U(s)$ (o sea, el conjunto de puntos $(\theta_Q^* - \theta)\sqrt{n}$; véase (11)) *). Entonces,

$$\{M_n \subset D\} = \left\{\min_{s\in D} U(s) < \min_{s\notin D} U(s)\right. . \tag{18}$$

Por consiguiente,

$$\{\sqrt{n}|\theta_Q^* - \theta| > 2N\} = \left\{\min_{|s|>2N} U(s) < \min_{|s|\leqslant 2N} U(s)\right\} \subset \left\{\min_{|s|\geqslant 2N} U(s) < U(0)\right\}.$$

Aquí

$$\min_{|s|\geqslant 2N} U(s) \geqslant w_n \int_{|u|<N} q\left(\theta + \frac{u}{\sqrt{n}}\right) Z\left(\frac{u}{\sqrt{n}}\right) du = w_N(u_1(0) - u_1(N)),$$

$$w_N \equiv \min_{\substack{|s|>2N \\ |u|<N}} w(s-u) = \min_{|t|>N} w(t).$$

Seguidamente,

$$U(0) = \int w(-u) q\left(\theta + \frac{u}{\sqrt{n}}\right) Z\left(\frac{n}{\sqrt{n}}\right) du \leqslant (u_1(0) - u_1(M) W_M + u(M),$$

donde $W_M = \max_{|t|\leqslant M} w(t)$,

De aquí obtenemos

$$\{\sqrt{n}|\theta_Q^* - \theta| > 2N\} \subset \{w_N(u_1(0) - u_1(N)) < w_M(u_1(0) - u_1(M)) + u(M)\} \subset$$
$$\subset \left\{\left(\frac{w_N}{W_M} - 1\right) u_1(0) < \frac{u(M)}{W_M} + \frac{u_1(N) w_N}{W_M} + u_1(M)\right\}.$$

En virtud de la condición 5) escojamos $M = \gamma N$, $\gamma < 1$ de modo que $w_n > 2W_M$ para todos los valores de N bastante grandes. Además, hagamos uso de las desigualdades $W_M > 2$ (para valores de M bastante grandes) $w_n < w(N) < e^{cN}$. Entonces es evidente que

$$\{\sqrt{n}|\theta_Q^* - \theta| > 2N\} \subset \{u_1(0) < u(\gamma N) + u_1(N)e^{cN}\}. \tag{19}$$

En virtud del lema 1 hallamos

$$\mathbf{P}_\theta\left(u(\gamma N) > \frac{1}{2} e^{-\alpha N^2}\right) \leqslant 2ae^{-\beta N^2\gamma^2 + \alpha N^2},$$

$$\mathbf{P}_\theta\left(u_1(N) > \frac{1}{2} e^{-cN - \alpha N^2}\right) \leqslant 2ae^{-\beta N^2 + \alpha N^2 + cN}.$$

---

*) En vez de M_n se podría examinar, por ejemplo, el menor punto (según la norma) en el que se alcanza mín $U(s)$.

Escogiendo $\alpha < \frac{1}{2}\beta\gamma^2$ obtenemos que, para valores de N bastante grandes, de (19) resulta

$$\mathbf{P}_\theta(\sqrt{n}\,|\theta_Q^* - \theta| > 2N) \leqslant 4ae^{-\alpha N^2} + \mathbf{P}_\theta(u_1(0) < e^{-\alpha N^2}).$$

Sólo nos queda hacer uso del lema 2, en virtud del cual

$$\mathbf{P}_\theta(u_1(0) < e^{-\alpha N^2}) \leqslant be^{-2\alpha N^2}.$$

El teorema queda demostrado.

§ 7. Criterios estadísticos óptimos para una función de pérdidas arbitraria. Criterio de la relación de verosimilitud como decisión asintóticamente bayesiana

1. Propiedades de optimización de los criterios estadísticos para una función de pérdidas arbitraria. En los párrafos precedentes hemos visto que muchos resultados principales de la teoría de estimación conservan su validez cualitativa al pasar a problemas más generales de la decisión estadística con pérdidas $w(\delta, \theta)$, $\delta \in D \subset R^k$, $\theta \in \Theta \subset R^k$, distintas de las cuadráticas.

El mismo cuadro se observa también en la teoría de verificación de las hipótesis. En el § 4 hemos visto que las reglas de decisión óptimas para los juegos con conjuntos finitos D y Θ y con función de pérdidas arbitraria, tienen la misma forma que los criterios óptimos para verificar un número finito de hipótesis simples, examinados en el § 3.1. Los resultados de los §§ 3.5—3.7, 3.9, 3.11, 3.13—3.15 también conservarán, en lo fundamental, su validez. En particular, los teoremas de los c.u.m.p., enunciados en los §§ 3.5—3.7, se transformarán en afirmaciones de las estrategias uniformemente mejores en los juegos estadísticos correspondientes ($\Theta \subset R^k$, $D = \{\delta_1, \delta_2\}$ es bipuntual), en los cuales, sin embargo, la función de pérdidas $w(\delta_i, \theta) = w_i(\theta)$, $w_i(\theta) = 0$ para $\theta \in \Theta_i$, $i = 1, 2$ ya no será obligatoriamente estadística ($w_i(\theta) = 1$ para $\theta \notin \Theta_i$), sino que tan sólo satisfará ciertas condiciones muy generales (por ejemplo, las propiedades de crecimiento monótono de $w_i(\theta)$ al alejarse θ de Θ_i). El papel de las clases K_ε, en las que hemos buscado los c.u.m.p., lo desempeñarán las clases de funciones de decisión $\pi(X)$, con valor máximo fijo ε de las "pérdidas de primer género":

$$\varepsilon = \sup_{\theta \in \Theta_1} W(\pi(\cdot), \theta) = \sup_{\theta \in \Theta_1} w_2(\theta)\mathbf{M}_\theta \pi(X, \delta_2). \tag{1}$$

Se minimizará el valor de las "pérdidas de segundo género":

$$W(\pi(\cdot), \theta) = w_1(\theta)\mathbf{M}_\theta \pi(X, \delta_1) \text{ para } \theta \in \Theta_2. \tag{2}$$

Aquí $\pi(X, \delta_i)$ significa la probabilidad de tomar la decisión δ_i a base del criterio π. Para abreviar la notación, pongamos, siguiendo el capítulo 3, $\pi(X, \delta_2) = \pi(X)$, así que $\pi(X, \delta_1) = 1 - \pi(X)$. La designación del criterio y del número $\pi(X, \delta_2)$ con ayuda de un solo símbolo $\pi(X)$ es cómoda y, como hemos visto antes, no produce equivocaciones.

En (1) y (2) se buscan los extremos de las expresiones que se distinguen de las expresiones correspondientes para las funciones estadísticas de pérdidas, tan sólo por los factores que no dependen de $\pi(X)$. Si estos factores poseen la propiedad natural de monotonía, entonces, al pasar al problema definido por (1) y (2), la exposición de los §§ 3.5—3.7, 3.9, 3.11 no variará considerablemente.

De hecho, también variarán poco los resultados de carácter asintótico en los §§ 3.13—3.15. En este párrafo examinaremos más detalladamente la generalización para el caso de una función de pérdidas arbitraria de los resultados del § 3.13 y nos convenceremos de que esta generalización realmente no exige ningunos esfuerzos adicionales.

2. C.r.v. como criterio asintóticamente bayesiano. Examinemos el juego estadístico $(\mathscr{D}, \Theta, W)$ en el que Θ es continual y constituye un conjunto compacto convexo en R^k, mientras que el conjunto D de estrategias del estadista es bipuntual: $D = \{\delta_1, \delta_2\}$. La función de pérdidas $w(\delta, \theta)$ tiene la forma siguiente:

$$w(\delta_1, \theta) = \begin{cases} 0, & \theta = \theta_1, \\ w_1(\theta), & \theta \neq \theta_1, \end{cases}$$

$$w(\delta_2, \theta) = \begin{cases} w_2, & \theta = \theta_1, \\ 0, & \theta \neq \theta_1, \end{cases}$$

donde θ_1 es un punto interior fijo de Θ. Cuando $w_2 = w_1(\theta) \equiv 1$ esto corresponde al problema de verificación de la hipótesis simple $H_1 = \{\theta = \theta_1\}$ frente a la alternativa adicional $H_2 = \{\theta \neq \theta_1\}$.

Para hallar, utilizando el principio bayesiano, la forma de decisión bayesiana, examinemos el juego corriente (D, Θ, w) y supongamos que en Θ se da una distribución $\mathbf{Q}$ tal, que $q = \mathbf{Q}(\{\theta_1\}) > 0$ (planteamiento bayesiano completo del problema). Designemos $\mathbf{Q}_2 = \dfrac{\mathbf{Q} - q\mathbf{I}_{\theta_1}}{1 - q}$, donde $\mathbf{I}_\theta$ es una distribución degenerada concentrada en el punto θ. Entonces

$$\tilde{w}(\delta_1, \mathbf{Q}) = (1 - q) \int w_1(t)\mathbf{Q}_2(dt), \quad \tilde{w}(\delta_2, \mathbf{Q}) = qw_2.$$

Esto quiere decir que la estrategia bayesiana $\pi_Q(\delta_2) = 1$ si

$$(1 - q) \int w_1(t)\mathbf{Q}_2(dt) > qw_2, \tag{3}$$

y $\pi_Q(\delta_1) = 1$ si tiene lugar la desigualdad inversa. La relación (3) puede escribirse en la forma

$$\int w(t)\mathbf{Q}(dt) > 0,$$

donde

$$w(t) = \begin{cases} w_1(t) & \text{para} \quad t \neq \theta_1, \\ -w_2 & \text{para} \quad t = \theta_1. \end{cases}$$

En virtud del principio bayesiano, la regla bayesiana de decisión $\pi_Q(X)$ tiene la forma $\pi_Q(X) = 1$ si

$$\int w(t)\mathbf{Q}_X(dt) > 0,$$

donde $\mathbf{Q}_x$ es la distribución a posteriori. Supongamos que $\lambda(dt) = dt$ para $t \neq \theta_1$, $\lambda(\{\theta_1\}) = 1$, y que la distribución $\mathbf{Q}_2$ tiene una densidad $q_2(t)$ respecto a la medida de Lebesgue. Entonces, la distribución $\mathbf{Q}$ tendrá una densidad $q(t)$ respecto a λ, igual a $(1 - q)q_2(t)$ para $t \neq \theta_1$, e igual a $q(t) = q$ para $t = \theta_1$. Esto significa que la densidad a posteriori respecto a la medida λ será igual a

$$q(t/X) = \frac{f_t(X)q(t)}{f(X)},$$

$$f(X) = \int f_u(X)q(u)\lambda(du).$$

Por consiguiente, la regla bayesiana de decisión $\pi_Q(X)$ tiene la forma $\pi_Q(X) = 1$ si

$$(1 - q)\int w_1(t)f_t(X)q_2(t)dt > w_2 q f_{\theta_1}(X). \tag{4}$$

El riesgo de esta regla es igual a

$$\tilde{W}(\pi_Q(\cdot), \mathbf{Q}) = q w_2 \mathbf{P}_{\theta_1}(\pi_Q(X) = 1) + \\ + (1 - q)\int w_1(u)q_2(u)\mathbf{P}_u(\pi_Q(X) = 0)du.$$

Comparando estas relaciones con el contenido del § 3.13, vemos que la región (4) de toma de decisión δ_2 tiene aquí la misma forma que la región $\Omega(c)$ en (3.13.3) cuando $c = w_2 q/(1 - q)$ y cuando la función $q(t)$ en (3.13.3) se sustituye por $w_1(t)q_2(t)$. En otros términos,

$$\pi_Q(X) = \begin{cases} 1, & \text{si} \quad r_{Q_2}(X) > c, \\ \gamma, & \text{si} \quad r_{Q_2}(X) = c, \\ 0, & \text{si} \quad r_{Q_2}(X) < c, \end{cases} \tag{5}$$

donde

$$r_{Q_2}(X) = \frac{\int w_1(t)q_2(t)f_t(X)dt}{f_{\theta_1}(X)}, \quad c = \frac{w_2 q}{1 - q}.$$

Luego, siguiendo los razonamientos del § 3.13, podemos proceder del modo siguiente. De la población de reglas bayesianas (5) es necesario, modificando el número q, elegir tal decisión $\pi_Q(X)$, que tenga un valor fijo de "pérdidas de primer género":

$$w_2[\mathbf{P}_{\theta_1}(\pi_Q(X) = 1) + \gamma\mathbf{P}_{\theta_1}(\pi_Q(X) = \gamma)] = \alpha.$$

Entonces, entre todas las reglas $\pi(X)$, para las cuales

$$\alpha_1(\pi) = w_2 \mathbf{M}_{\theta_1}\pi(X) \leqslant \alpha, \tag{6}$$

la decisión $\pi_Q(X)$ minimizará las "pérdidas de segundo género" iguales a

$$\alpha_2(\pi) = \int w_1(u)q_2(u)\mathbf{M}_u(1 - \pi(X))du. \tag{7}$$

Esto es la consecuencia directa del carácter bayesiano de la decisión π_Q. La comparación de los valores (6) y (7) con las magnitudes de las probabilidades de los errores de primero y segundo géneros (3.13.4.) muestra que otra vez se trata de distinciones no esenciales, la principal de las cuales consiste en que la función $q(u)$ en (3.13.4.) se sustituye por la función $w_1(u)q_2(u)$. Los números c y γ en (5) se determinan por α.

Lo dicho nos permite, siguiendo exactamente los razonamientos del § 3.13, enunciar las siguientes definiciones y afirmaciones.

Definición 1. La regla de decisión $\pi(X)$ pertenece a la clase $\tilde{\mathbf{K}}_\varepsilon$ (su nivel asintótico es $1 - \varepsilon$) si

$$\operatorname*{lím\,sup}_{n\to\infty} \mathbf{M}_{\theta_1}\pi(X) \leqslant \varepsilon.$$

Esta definición, de hecho, no se diferencia en nada de la definición 3.13.1.

Mostremos ahora que, eligiendo q, podemos tratar de que $\pi_Q \in \tilde{\mathbf{K}}_\varepsilon$. Pongamos

$$r_{Q_1}(X) = \frac{\int w_1(t)q_2(t)f_t(X)dt}{f_{\theta_1}(X)} = \left(\frac{2\pi}{n}\right)^{k/2} \frac{w_1(\theta_1)q_2(\theta_1)}{\sqrt{|I|}}\, e^{T(X)},$$

donde $I = I(\theta_1)$ es la matriz de información de Fisher en el punto θ_1. Supongamos, seguidamente, que se cumplen las condiciones (RR), que θ_1 es un punto interior en Θ, y que la función $w_1(t)\cdot q_2(t)$ es continua y positiva en el punto θ_1,

$$c = \left(\frac{2\pi}{n}\right)^{k/2} \frac{w_1(\theta_1)q_2(\theta_1)}{\sqrt{|I|}}\, e^{z}. \tag{8}$$

Entonces, en viritud del lema 3.13.1., para la función $p_t(c) = \mathbf{P}_t(r_{Q_1}(X) > c)$ obtenemos

$$p_{\theta_1}(c) = \mathbf{P}_{\theta_1}(T(X) > z) \to \mathbf{H}_k((2z,\ \infty)).$$

Por consiguiente, poniendo $q = c/(c + w_2)$, donde c está definida en (8), $z = h_\varepsilon/2$, h_ε es una cuantila de orden $1 - \varepsilon$ de la distribución χ^2 con k grados de libertad, obtenemos

$$\lim_{n\to\infty} p_{\theta_1}\left(\frac{w_2 q}{1 - q}\right) = \varepsilon$$

y, por lo tanto, $\pi_Q(X) \in \tilde{\mathbf{K}}$.

Definición 2. Para una distribución a priori dada $\mathbf{Q}$, la regla de decisión $\pi(X)$ se llama *asintóticamente bayesiana* en $\tilde{\mathbf{K}}_\varepsilon$ si $\pi_Q \in \tilde{\mathbf{K}}_\varepsilon$,

$$\lim_{n\to\infty} \sup \frac{\alpha_2(\pi)}{\alpha_2(\pi_Q)} = 1.$$

Teorema 1. *Supongamos que se cumplen las condiciones (RR) y que θ_1 es un punto interior en θ. Entonces, en $\tilde{\mathbf{K}}_\varepsilon$ existe una regla de decisión asintóticamente bayesiana $\hat{\pi}(X)$ que es la misma para cualesquiera distribuciones $\mathbf{Q}_2$ y para cualesquiera funciones $w_1(t)$ tales, que la función $w_1(t)q_2(t)$ es continua y positiva en el punto θ_1 y está limitada en Θ. El criterio π es definido por la relación*

$$\hat{\pi}(X) = 1 \quad \text{si} \quad \frac{f_{\theta^*}(X)}{t_{\theta_1}(X)} > e^{h_\varepsilon/2}. \tag{9}$$

El teorema se demuestra exactamente igual que el teorema 3.13.1, con una precisión de hasta la sustitución de la función $q(t)$ por $w_1(t)q_2(t)$. El teorema 3.13.1 también permite hallar el valor de las "pérdidas de segundo género" (véase (7)) del criterio $\hat{\pi}$.

El criterio (9) no es otra cosa sino el criterio de relación de verosimilitud.

§ 8. Soluciones asintóticamente óptimas para una función de pérdidas arbitrarias en el caso de hipótesis semejantes

En este párrafo examinaremos la generalización de los resultados del § 3.14 para el caso de una función de pérdidas arbitrarias. Esta generalización será más sustancial que en el párrafo anterior, ya que las funciones de pérdidas dependerán de n (compárese con el § 6).

Supongamos que $(\mathscr{D}, \Theta, W)$ es un juego estadístico en el que $\Theta \subset R^k$, el conjunto $D = \{\delta_1, \delta_2\}$ es bipuntual y $w(\delta_i, \theta) = w_i(\theta)$, donde $w_i(\theta) = 0$ cuando $\theta \in \Theta_i$, $i = 1, 2$, y la intersección $\Theta_1 \cap \Theta_2$ está vacía.

Si $w_i(\theta) = 1$ cuando $\theta \notin \Theta_i$, obtendremos el problema de verificación de las hipótesis $H_i = \{\theta \in \Theta_i\}$, $i = 1, 2$.

Determinemos la estrategia bayesiana para el juego (D, Θ, w). Sean $\mathbf{Q}_i$ las distribuciones en Θ_i,

$$\mathbf{Q} = q_1\mathbf{Q}_1 + q_2\mathbf{Q}_2, \quad q_1 + q_2 = 1.$$

Entonces es evidente que $\tilde{w}(\delta_i, \mathbf{Q}) = \int w_i(t)\mathbf{Q}(dt)$ y $\pi_Q(\delta_2) = 1$ si

$$\int w_2(t)\mathbf{Q}(dt) < \int w_1(t)\mathbf{Q}(dt),$$

o bien

$$q_1\int w_2(t)\mathbf{Q}_1(dt) < q_2\int w_1(t)\mathbf{Q}_2(dt).$$

Por consiguiente, en virtud del principio bayesiano, la regla bayesiana de decisión $\pi_Q(X)$ tendrá la forma $\pi_Q(X) = 1$ si

$$\int w_2(t)\mathbf{Q}_X(dt) < \int w_1(t)\mathbf{Q}_X(dt). \tag{1}$$

Supongamos que las distribuciones $\mathbf{Q}_i$ tienen densidades $q_i(t)$, $i = 1$, 2 respecto a la medida λ. Entonces, $\mathbf{Q}$ y la distribución a posteriori $\mathbf{Q}_x$ tendrán, respectivamente, densidades $q(t) = q_1q_1(t) + q_2q_2(t)$ y

$$q(t/X) = \frac{q(t)f_t(X)}{f(X)}, \quad f(X) = \int q(u)f_u(X)\lambda(du).$$

Esto significa que la relación (1) se puede escribir en la forma

$$q_1\int_{\Theta_1} w_2(t)q_1(t)f_t(X)\lambda(dt) < q_2\int_{\Theta_2} w_1(t)q_2(t)f_t(X)\lambda(dt). \tag{2}$$

El riesgo de la regla bayesiana $\pi_Q(X)$ es igual a

$$W(\pi_Q(\cdot), \theta) = w_1(\theta)\mathbf{M}_\theta\pi_Q(X) + w_2(\theta)(1 - \mathbf{M}_\theta\pi_Q(X)),$$

$$\bar{W}(\pi_Q(\cdot), \mathbf{Q}) = \int W(\pi_Q(\cdot), t)q(t)\lambda(dt).$$

Pasemos ahora a examinar las *alternativas semejantes.* Sea θ_1 cualquier valor fijo del parámetro θ. Al igual que en el § 3.14 supondremos que los conjuntos Θ_i tienen la forma siguiente:

$$\Theta_i = \theta_1 + \Gamma_i/\sqrt{n}, \tag{3}$$

donde Γ_i no depende de n. En lo que se refiere a Q_i, supondremos que éstas están inducidas por ciertas distribuciones Π_i concentradas en Γ_i y que no dependen de n. Si los conjuntos Γ_i están limitados, entonces, las estrategias de naturaleza θ estarán situadas en el $1/\sqrt{n}$-entorno del punto θ_1. Por eso, si $w_1(t)$, $w_2(t)$ son continuas y $w_i(t) > c > 0$, $i = 1, 2$ en los conjuntos Θ_2 y Θ_1, respectivamente, entonces, el juego estadístico $(\mathscr{D}, \Theta, W)$ para tal función de pérdidas no se distinguirá (según sus propiedades) del juego cuya función estadística de pérdidas constituye $w_i(t) = 1$ para $t \notin \Theta_i$ examinado en los §§ 3.14 y 3.15.

Aquí examinaremos una generalización más sustancial, análoga a la ejecutada en el § 6. Supondremos que la función de pérdidas $w(\delta_i, \theta) = w_i(\theta)$ depende de n de tal modo que

$$w_i(\theta) = w_{i,n}(\theta) = v_i(\sqrt{n}(\theta - \theta_1)), \tag{4}$$

donde $v_i(t)$ son funciones medibles limitadas que no dependen de n.

Siguiendo el § 3.14, llamaremos *problema* A al problema de búsqueda de la solución del juego $(\mathscr{D}, \Theta, W)$, descrito anteriormente, con ayuda de la muestra $X \in \mathbf{P}$). Si se cumplen (3) y (4), hablaremos del problema A para hipótesis semejantes, con funciones de pérdidas $v_i(t)$.

Examinemos ahora otro juego estadístico $(\mathscr{D}_B, \Gamma, V)$ referente a la muestra $V \in \Phi_{\nu, I^{-1}}$ de volumen unitario, donde $I = I(\theta_1)$ es la matriz de información de Fisher para la familia $\mathbf{P}_\theta$ en el punto θ_1. Este juego tiene el conjunto bipuntual de soluciones $D_B = (d_1, d_2)$ y el conjunto de estrategias de naturaleza (conjunto paramétrico) $\Gamma = \Gamma_1 \cup \Gamma_2$. La función de pérdidas $v(d, \gamma)$: $D_B \times \Gamma \to R$ se define por las relaciones

$$v(d_i, \gamma) = v_i(\gamma), \quad v_i(\gamma) = 0 \quad \text{para} \quad \gamma \in \Gamma_i.$$

Ahora bien, en este juego, $\mathscr{D}_B$ es la clase de todas las soluciones $d(Y)$: $\mathscr{Y} = R^k \to D_B$,

$$V(d(\cdot), \gamma) = v_1(\gamma)\Phi_{\nu, I^{-1}}(d(Y) - d_1) + v_2(\gamma)\Phi_{\nu, I^{-1}}(d(Y) = d_2)$$

(uno de los sumandos del segundo miembro es igual a cero). Análogamente se escriben las pérdidas para las estrategias randomizadas $\pi(Y)$ en los términos $\mathbf{M}\pi(Y)$, $Y \in \Phi_{\nu, I^{-1}}$. Llamaremos *problema* B al problema enunciado.

Entre los problemas A y B aquí existe la misma relación que fue establecida entre estos problemas en el § 3.14. Sea $\pi(Y)$ la solución del problema B, óptima en uno u otro sentido (bayesiana o minimax). Y sea $\hat{\theta}^*$ la e.v.m. en el problema A, $\gamma^* = (\hat{\theta}^* - \theta_1)\sqrt{n}$. Entonces, $\pi(\gamma^*)$ será la solución asintóticamente óptima del problema A (en ese mismo sentido).

El "criterio límite de optimización" permite reducir el problema A a un problema más simple B.

Para que lo dicho adquiera sentido exacto daremos las definiciones siguientes. Supongamos que en Γ_i se dan las distribuciones $\mathbf{\Pi}_i$. Pongamos $\mathbf{\Pi} = q_1\mathbf{\Pi}_1 + q_2\mathbf{\Pi}_2$, $q_1 + q_2 = 1$ y designemos por $\mathbf{Q}$ la distribución en Θ, inducida por la distribución $\mathbf{\Pi}$ y por la transformación $\theta = \theta_1 + \gamma/\sqrt{n}$.

Definición 1. La solución $\pi_1(X)$ se llama *asintóticamente bayesiana* si

$$\lim_{n\to\infty} \sup\, [\tilde{W}(\pi_1(\cdot), \mathbf{Q}) = \tilde{W}(\pi_Q(\cdot), \mathbf{Q})] \leqslant 0.$$

Aquí, al igual que antes,

$$\tilde{W}(\pi(\cdot), \theta) = w_1(\theta)\mathbf{M}_\theta\pi(X) + w_2(\theta)(1 - \mathbf{M}_\theta\pi(X)),$$

$$\tilde{W}(\pi(\cdot), \theta) = \int \tilde{W}(\pi(\cdot), t)\mathbf{Q}(dt),$$

π_Q es la regla de decisión bayesiana.

Definición 2. La solución $\pi_1(X)$ se denomina *asintóticamente minimax* si para cualquier otra solución $\pi(X)$

$$\lim_{n\to\infty} [\sup_{\theta\in\Theta} \sup \tilde{W}(\pi_1(\cdot), \theta) - \sup_{\theta\in\Theta} \tilde{W}(\pi(\cdot), \theta)] \leqslant 0.$$

Aquí se podría comparar π_1 sólo con la regla minimax $\bar{\pi}$ (compárese con la definición 1).

Análogamente al § 3.14 también podríamos examinar las soluciones asintóticamente bayesianas y minimax en la clase $\tilde{\mathbf{K}}_\varepsilon$ de soluciones de las "pérdidas de primer género" asintóticas fijas:

$$\varepsilon = \lim_{n\to\infty} \sup \sup_{\theta\in\Theta_1} w_1(\theta)\mathbf{M}_\theta\pi(X).$$

Para obtener los resultados respectivos es suficiente comparar el contenido de este párrafo con el del § 3.14.

Designemos por $\pi_\Pi(Y)$ la solución bayesiana del juego $(\mathscr{D}_B, \Gamma, V)$ (o sea, del problema B), la cual corresponde a la distribución a priori $\mathbf{\Pi}$, y supongamos, para abreviar, que los conjuntos Γ_i están limitados.

Teorema 1. *Supongamos que en el entorno del punto θ_1 se cumplen las condiciones (RR), y que las funciones v_i y la distribución $\mathbf{\Pi}_i$ son tales que $0 < \int v_1(u)\mathbf{\Pi}_2(du) < \infty$, $0 < \int v_2(u)\mathbf{\Pi}_1(du) < \infty$. Entonces, en las designaciones introducidas, el criterio*

$$\pi_1(X) = \pi_\Pi(\gamma^*), \quad \gamma^* = (\hat{\theta}^* - \theta_1)\sqrt{n}$$

será la solución asintóticamente bayesiana del juego $(\mathscr{D}, \Theta, W)$ (o sea, del problema A*), la cual corresponde a la distribución a priori* $\mathbf{Q}$.

Teorema 2. *Supongamos que en el entorno del punto θ_1 se cumplen las condiciones* (*RR*) *y que en el problema B existe la solución minimax $\bar{\pi}(Y)$ y la peor distribución correspondiente $\overline{\mathbf{\Pi}}$. Entonces, el criterio $\pi_1(X) = \bar{\pi}(\gamma^*)$ será la solución asintóticamente minimax del problema* A.

Observación 1. Las condiciones del teorema de la existencia de $\bar{\pi}$ y $\overline{\mathbf{\Pi}}$ en virtud de los teoremas del § 3, serán cumplidas siempre que v_i sean funciones continuas.

La demostración del teorema 1 es completamente análoga a la del teorema 3.14.1. De (2) se deduce que la regla bayesiana de decisión π_Q tendrá la forma $\pi_Q(X) = 1$ si

$$\frac{\int w_1(t)q_2(t)f_t(X)\lambda(dt)}{\int w_2(t)q_1(t)f_t(X)\lambda(dt)} > \frac{q_1}{q_2}. \tag{5}$$

Poniendo $Z_1(t) = \dfrac{f_{\theta_1+t}(X)}{t_{\theta_1}(X)}$ y teniendo en cuenta que

$$q_i(t)\lambda(dt) = \mathbf{Q}_i(dt), \quad \mathbf{Q}_i(\theta_1 + du/\sqrt{n}) = \mathbf{\Pi}_i(du),$$
$$w_i(\theta_1 + u/\sqrt{n}) = v_i(u),$$

con ayuda de la sustitución de $t = \theta_1 + u/\sqrt{n}$ podemos transformar la desigualdad (5) reduciéndola a la forma

$$\frac{\int v_1(u)Z_1(u/\sqrt{n})\mathbf{\Pi}_2(du)}{\int v_2(u)Z_1(u/\sqrt{n})\mathbf{\Pi}_1(du)} = \frac{\int Z_1(u/\sqrt{n})\mathbf{\Pi}_2'(du)}{\int Z_1(u/\sqrt{n})\mathbf{\Pi}_1'(du)} > c, \quad c = \frac{q_1}{q_2}, \tag{6}$$

donde las distribuciones generalizadas $\mathbf{\Pi}_i'(A) = \int_A v_{i+1}(u)\mathbf{\Pi}_i(du)$ $(v_3(u) \equiv \equiv v_1(u),\ i = 1, 2)$ pueden ser transformadas, mediante renormalización, en probabilísticas, introduciendo las transformaciones $\mathbf{\Pi}_i''(A) = \mathbf{\Pi}_i'(A)/\mathbf{\Pi}_i'(\Gamma_i)$ (según las condiciones $0 < \mathbf{\Pi}_i'(\Gamma_i) < \infty$). Entonces, en calidad de (5) obtendremos la desigualdad que tiene exactamente la misma forma que en el § 3.14.

Los razonamientos ulteriores de la demostración se distinguen de los razonamientos respectivos del § 3.14 tan sólo por las simplificaciones. Esta tarea se la dejamos a cargo del lector. Dichos razonamientos se basan en la convergencia uniforme de $(\theta = \theta_1 + \gamma/\sqrt{n})$ en γ:

$$\tilde{W}(\pi_Q(\cdot),\ \theta) \to \tilde{V}(\pi_{\Pi}(\cdot),\ \gamma), \quad \tilde{W}(\pi_1(\cdot),\ \theta) \to \tilde{V}(\pi_{\Pi}(\cdot),\ \gamma), \tag{7}$$

donde $\pi_1(X) = \pi_{\Pi}(\gamma^*)$. ◁

Para demostrar el teorema 2 necesitaremos el

Lema 1. *Sea* $\mathbf{Q}$ *la distribución a priori, y* π_1, *la solución asintóticamente bayesiana que le corresponde, tal que*

$$\varlimsup_{n\to\infty} \tilde{W}(\pi_1(\cdot),\ \mathbf{Q}) = c, \quad \varlimsup_{n\to\infty} \sup_{\theta\in\Theta} \tilde{W}(\pi_1(\cdot),\ \theta) \leqslant c. \tag{8}$$

Entonces, π_1 *es la solución asintóticamente minimax.*

Demostración. Al igual que antes, designemos por π_Q la solución bayesiana. Entonces, para cualquier solución π tenemos

$$\varlimsup_{n\to\infty} \sup_{\theta\in\Theta} \tilde{W}(\pi,\ \theta) \geqslant \varlimsup_{n\to\infty} \tilde{W}(\pi,\ \mathbf{Q}) \geqslant$$

$$\geqslant \varlimsup_{n\to\infty} \tilde{W}(\pi_Q,\ \mathbf{Q}) \geqslant \varlimsup_{n\to\infty} \tilde{W}(\pi_1,\ \mathbf{Q}) =$$

$$= c \geqslant \varlimsup_{n\to\infty} \sup_{\theta\in\Theta} \tilde{W}(\pi_1,\ \theta). \ ◁$$

Demostración del teorema 2. Sea $\overline{\mathbf{\Pi}}$ la peor distribución en Γ, de modo que $\bar{\pi}(Y) = \pi_{\overline{\Pi}}(Y)$ sea la regla minimáx de decisión en el juego $(\mathscr{D}_B,\ \Gamma,\ V)$. Entonces, según el teorema 1, $\pi_1(X) = \pi_{\overline{\Pi}}(\gamma^*)$ será la solución asintóticamente bayesiana para la distribución $\overline{\mathbf{Q}}$ que corresponde a $\overline{\mathbf{\Pi}}$, y para demostrar el teorema nos es suficiente convencernos que $\overline{\mathbf{Q}}$ y π_1 satisfacen las condiciones del lema 1.

Designemos por N_{Π} el portador de la distribución $\overline{\mathbf{\Pi}}$. Entonces, en virtud de los teoremas del § 3.

$$\tilde{V}(\pi_{\tilde{\Pi}}(\cdot),\ \gamma) = c \quad \text{para} \quad \gamma \in \mathrm{N}_{\Pi},$$
$$\sup_{\gamma \in \Gamma} \tilde{V}(\pi_{\tilde{\Pi}}(\cdot),\ \gamma) \leqslant c. \tag{9}$$

Pero para $\theta = \theta_1 + \gamma/\sqrt{n}$ tiene lugar (véase (7)) la convergencia $\tilde{W}(\pi_1(\cdot), \theta) \to \tilde{V}(\pi_{\tilde{\Pi}}(\cdot),\ \gamma)$ uniforme en γ. De aquí y de (9) resulta (8). El teorema queda demostrado.

Suplemento I

Teoremas del tipo de Glivenko — Cantelli

En este Suplemento demostraremos las afirmaciones a base de las cuales se deducirán los teoremas 1.4.1. y 1.4.2. Utilizaremos, sin aclaraciones, las designaciones del párrafo 1.4 en el que estos teoremas han sido enunciados. Primero demostraremos la variante general auxiliar del teorema de Glivenko — Cantelli.

Definición 1. Llamaremos *aproximable finita* (respecto a la distribución **P**) la clase $\mathfrak{R}$ de conjuntos de $\mathfrak{B}_{\mathscr{X}} = \mathfrak{B}^m$, si cualquiera que sea $\varepsilon > 0$, para éste existe otra clase de conjuntos $\mathfrak{S}(\varepsilon)$, constituída por un número finito $N = N(\varepsilon)$ de elementos $S_1, \ldots, S_N$, $S_i \in \mathfrak{B}^m$, tal que para cualquier $B \in \mathfrak{R}$ habrá conjuntos A_1 y A_2 de $\mathfrak{S}(\varepsilon)$ dotados de las propiedades siguientes:

$$A_1 \subset B \subset A_2,$$
$$\mathbf{P}(A_2 - A_1) < \varepsilon. \tag{1}$$

Definamos sobre las clases de conjuntos, las operaciones de adición, de multiplicación y de complemento. Denominaremos clases $\mathfrak{R}_1 + \mathfrak{R}_2$ y $\mathfrak{R}_1\mathfrak{R}_2$ las clases de conjuntos del tipo $A \cup B$ y $A \cap B$, respectivamente, donde $A \in \mathfrak{R}_1$, $B \in \mathfrak{R}_2$. Llamaremos complemento $\overline{\mathfrak{R}}$ la clase de conjuntos formada por los complementos $\overline{A}$, $A \in \mathfrak{R}$.

Teorema 1. 1) *Supongamos que $X_n = [X_\infty]_n$, $X_\infty \in \mathbf{P}$ y que la clase $\mathfrak{R}$ es aproximable finita. Entonces*

$$\sup_{B\in\mathfrak{R}} |\mathbf{P}_n^*(B) - \mathbf{P}(B)| \underset{c.s}{\to} 0. \tag{2}$$

2) *La población de clases aproximables finitas está cerrada respecto a las operaciones introducidas.*

Demostración. La primera afirmación se obtiene con las mismas consideraciones que hemos usado en el caso unidimensional del teorema 1.22. Para los valores dados de $B \in \mathfrak{R}$ y $\varepsilon > 0$ existen $N = N(\varepsilon)$ y conjuntos A_1, A_2 dotados de la propiedad (1). Para ellos tenemos

$$\mathbf{P}_n^*(B) - \mathbf{P}(B) \leqslant \mathbf{P}_n^*(A_2) - \mathbf{P}(A_1) < \mathbf{P}_n^*(A_2) - \mathbf{P}(A_2) + \varepsilon,$$
$$\mathbf{P}_n^*(B) - \mathbf{P}(B) \geqslant \mathbf{P}_n^*(A_1) - \mathbf{P}(A_2) > \mathbf{P}_n^*(A_1) - \mathbf{P}(A_1) - \varepsilon.$$

Por eso

$$\bigcap_{k=1}^{N} \{ |\mathbf{P}_n^*(S_k) - \mathbf{P}(S_k)| < \varepsilon\} \subset \{ \sup_{B\in\mathfrak{R}} |\mathbf{P}_n^*(B) - \mathbf{P}(B)| < 2\varepsilon\},$$

donde $S_1, \ldots, S_N$ son los elementos de $\mathfrak{S}(\varepsilon)$. Como $\mathbf{P}_n^*(S_k) \underset{c.s}{\to} \mathbf{P}(S_k)$, de aquí ya sin dificultad obtenemos (2) (compárese con la demostración del teorema 1.2.2.A).

La segunda afirmación del teorema 3 es casi evidente. Supongamos que tenemos $\varepsilon > 0$ y que $\mathfrak{S}_1(\varepsilon_1)$ y $\mathfrak{S}_2(\varepsilon_2)$ son las clases aproximantes para $\mathfrak{R}_1$ y $\mathfrak{R}_2$, respectivamente. Sean, además, A y B conjuntos cualesquiera de $\mathfrak{R}_1$ y de $\mathfrak{R}_2$. De las relaciones $\varepsilon_1 + \varepsilon_2 = \varepsilon$,

$$A_1 \subset A \subset A_2, \quad \mathbf{P}(A_2 - A_1) < \varepsilon_1 \quad (A_i \in \mathfrak{S}_1(\varepsilon_1)),$$
$$B_1 \subset B \subset B_2, \quad \mathbf{P}(B_2 - B_1) < \varepsilon_2 \quad (B_i \in \mathfrak{S}_2(\varepsilon_2)),$$

obtenemos

$$A_1B_1 \subset AB \subset A_2B_2,$$
$$A_2B_2 - A_1B_1 \subset (A_2 - A_1) \cup (B_2 - B_1),$$
$$\mathbf{P}(A_2B_2 - A_1B_1) \leqslant \varepsilon.$$

Por lo tanto, la clase $\mathfrak{R}_1\mathfrak{R}_2$ es aproximable finita. La suma $\mathfrak{R}_1 + \mathfrak{R}_2$ y el complemento $\overline{\mathfrak{R}}$ se examinan análogamente. ◁

Corolario 1. *Sea $\mathscr{X} = R^m$, $X_n = [X_\infty]_n \in F$. Entonces,*

$$\sup_t |F_n^*(t) - F(t)| \xrightarrow{c.s.} 0$$

cuando $n \to \infty$, donde $F_n^*(t)$ es la función empírica de distribución.

Demostración. De la demostración del teorema I.2.2A se deduce que las clases de subconjuntos $\mathfrak{R}_j = \{y \in R^m: y_j < t_j\}$, $-\infty < t_j < \infty$, para cada $j = 1, \ldots, m$, son clases aproximables finitas. En calidad del sistema $\mathfrak{S}(\varepsilon)$ es suficiente adoptar los semiespacios $\{y_j < z_k\}$ e $\{y_j \leqslant z_k\}$, $k = 1, \ldots, N$, donde z_k se han definido en (1.2.6).

Según la segunda afirmación del teorema 1, la clase de ángulos $\mathfrak{R} = \mathfrak{R}_1\mathfrak{R}_2 \ldots \mathfrak{R}_m$ también será aproximable finita. Nos queda hacer uso de la primera afirmación del teorema 1. ◁

El corolario 1 no es otra cosa sino el teorema 1.4.1..

Examinemos ahora las clases de conjuntos $\mathfrak{R}$ que satisfacen la condición siguiente (Γ). Sea K_M el cubo

$$K_M = \{y = (y_1, \ldots, y_m): \max_{1 \leqslant k \leqslant m} |y_k| \leqslant M\}.$$

(Γ) *Todos los conjuntos $B \in \mathfrak{R}$ poseen la siguiente propiedad: el ε-entorno Γ_B^ε de la frontera $\Gamma_B = \partial(B \cap K_M)$ tiene medida de Lebesgue (volumen) $\mu(\Gamma_B^\varepsilon \leqslant \varphi(\varepsilon, M)$, donde φ sólo depende de sus argumentos, y para cualquier M, $\varphi(\varepsilon, M) \to 0$ cuando $\varepsilon \to 0$.*

Teorema 2. *Supongamos que $\mathscr{X} = R^m$, $X \in P$ y la distribución P es absolutamente continua con respecto a la medida de Lebesgue. Entonces la clase $\mathfrak{R}$ que satisface la condición (Γ) es aproximable finita y, por consiguiente, para ella es válida* (2).

Demostración. Notemos antes que nada, que el problema cuyo espacio constituye R^m puede ser reducido al cubo K_M en el sentido siguiente. Supongamos que para cualquier M fija hay una clase $\mathfrak{S}$ de subconjuntos de K_M tal, que para cualquier $B' \in \mathfrak{R}$ y $B = B' \cap K_M$ se cumple (1). Entonces $\mathfrak{R}$ será aproximable finita. En efecto, para $\varepsilon > 0$, elegido en (1), hallemos $M = M(\varepsilon)$ tal, que $\mathbf{P}(K_M) \geqslant 1 - \varepsilon$, y pongamos $A_1' = A_1$, $A_2' = A_2 \cup \overline{K_M}$, donde A_i es un conjunto de (1), y $\overline{K_M}$ es el complemento hasta K_M. Entonces es evidente que

$$A_1' \subset B' \subset A_2', \quad \mathbf{P}(A_2' - A_1') \leqslant 2\varepsilon.$$

Así pues, podemos considerar que $\mathbf{P}(K_M) = 1$, $\mathfrak{R}$ consta de los sobconjuntos K_M.

Examinemos, en calidad de $\mathfrak{S}$, las figuras A_j formadas por distintas uniones de cubos cerrados, con aristas de longitud δ y con los vértices en los puntos

$$(j_1\delta, \ldots, j_m\delta), \; -M/\delta < j_k < M/\delta, \quad k = 1, \ldots, m,$$

(para abreviar se puede admitir que δ divide totalmente (M). Definamos los conjuntos A_1,

A_2, respectivamente, como las uniones de todos los cubos que pertenecen y rozan con B. Es evidente que

$$A_1 \subset B \subset A_2,$$
$$\mu(A_2 - A_1) \leqslant \mu(\Gamma_B^{2\delta\sqrt{m}}) \leqslant \varphi(2\delta\sqrt{m},\ M).$$

Eligiendo δ, el segundo miembro de esta desigualdad puede hacerse tan pequeño cuanto se quiera.

Seguidamente, $\mathbf{P}$ es en absoluto continua respecto a μ. Por eso, para ε dado se puede hallar $\gamma = \gamma(\varepsilon)$ tal, que $\sup\limits_{\mu(A)<\gamma} \mathbf{P}(A) < \varepsilon$. Ahora, si δ se elige de tal modo que $\varphi(2\delta\sqrt{m}, M) <$ $< \gamma$, entonces obtendremos

$$\mathbf{P}(A_2 - A_1) < \varepsilon. \ \lhd$$

Corolario 2. *La clase $\mathfrak{S}$ de todos los conjuntos convexos es aproximable finita y, por lo tanto, para* $\mathbf{P}$ *absolutamente continuas,*

$$\sup_{B\in\mathfrak{S}} |\mathbf{P}_n^*(B) - \mathbf{P}(B)| \xrightarrow[\text{c.s.}]{} 0.$$

En efecto, el "área" máxima de la superficie del conjunto convexo en K_M constituye $2m(2M)^{m-1}$ y equivale al "área" de la superficie K_M: y el volumen máximo $\mu(\partial K_M)^\varepsilon)$ del ε-entorno de ∂K_M no pasa de $2\varepsilon \cdot 2m(2M)^{m-1}$. Esto significa que se cumple la condición (Γ). $\lhd$

El corolario 2 coincide con el teorema 1.4.2. La observación en cuanto a la existencia de la condición de continuidad absoluta de $\mathbf{P}$ está presente en el § 1.4.

No es difícil notar que la condición (Γ) también será cumplida para las clases de conjuntos no convexos dotados de fronteras bastante suaves.

Suplemento II

Teorema funcional del límite para los procesos empíricos

Aquí demostraremos la afirmación siguiente (teorema 1.6.3). Sea

$$w^n(t) = \sqrt{n}(F_n^*(t) - t)$$

el proceso empírico definido en el § 1.6, y sea $w^0(t)$ el puente browniano.

Teorema 1. *Si f es una funcional medible: $D(0, 1) \to R$, continua en los puntos del espacio $C(0, 1)$ y en una métrica uniforme, entonces, cuando $n \to \infty$,*

$$f(w^n) \Rightarrow f(w^0).$$

Para demostrar el teorema necesitaremos dos lemas.

Lema 1. *Las distribuciones de dimensión finita de los procesos w^n convergen débilmente (cuando $n \to \infty$) hacia las distribuciones respectivas del proceso w^0.*

Demostración. Examinemos los vectores aleatorios de dimensión $(m + 1)$,

$$w^n = (\Delta_0 w^n,\ \ldots,\ \Delta_m w^n),$$

donde, al igual que en el § 1.6, Δ_{jg} designa las diferencias

$$\Delta_{jg} = g(t_{j+1} - g(t_j),$$
$$t_{j+1} > t_j,\quad j = 0,\ \ldots,\ m,\quad t_0 = 0,\quad t_{m+1} = 1.$$

Designemos por w^0 el vector análogo para el proceso $w^0(t)$. En virtud del segundo teorema de continuidad, para demostrar el lema es suficiente mostrar que $w^n \Rightarrow w^0$.

Hallemos las funciones características w^n y w^0. Para el vector $u = (u_0, \ldots, u_m)$ tenemos

$$\mathbf{M}e^{iw^0u^T} = \mathbf{M}\exp\left\{i\sum_{j=0}^{m} u_j\Delta_j w^0\right\} = \mathbf{M}\exp\left\{i\sum_{j=0}^{m} u_j(\Delta_j w - w(1)\Delta_j\right\},$$

donde $\Delta_j = t_{j+1} - t_j$, $j = 0, \ldots, m$, $w(t)$ es un proceso wieneriano estándar.

Representemos el exponente de la exponencial como una suma de magnitudes independientes. Para abreviar designemos $\sum_{j=0}^{m} u_j\Delta_j = U$, obtendremos

$$\sum_{j=0}^{m} u_j(\Delta_j w - w(1)\Delta_j) = \sum_{j=0}^{m} (u_j - U)\Delta_j w.$$

En vista de que $\mathbf{M}e^{iuw(s)} = e^{-u^2s/2}$, entonces

$$\mathbf{M}e^{iw^0u^T} = \exp\left\{-\frac{1}{2}\sum_{j=0}^{m}(u_j - U)^2\Delta_j\right\} = \exp\left\{-\frac{1}{2}\left(\sum_{j=0}^{m} u_j^2\Delta_j - U^2\right)\right\}. \tag{1}$$

Ahora examinemos $\mathbf{M}e^{iw^nu^T}$. Sea, al igual que antes (véase el § 1.6),

$$\pi_n(t) = nF_n^*(t).$$

Entonces, como ya sabemos (véase (1.6.1)),

$$\mathbf{P}(\Delta_0\pi_n = k_0, \ldots, \Delta_m\pi_n = k_m) = \frac{n!}{k_0! \ldots k_m!}\Delta_0^{k_0} \ldots \Delta_m^{k_m}.$$

En el segundo miembro figuran los términos del desarrollo del polinomio $(\Delta_0 + \ldots + \Delta_m)^n$. Utilizando este argumento obtenemos

$$\mathbf{M}e^{i\sum_{j=0}^{m} u_j\Delta_j\pi_n} = \sum \frac{n!}{k_0! \ldots k_m}(e^{iu_0}\Delta_0)^{k_0} \ldots (e^{iu_m}\Delta_m)^{k_m} = (e^{iu_0}\Delta_0 + \ldots + e^{iu_m}\Delta_m)^n.$$

Como $\Delta_j w^n = \sqrt{n}(F_n^*(t_{j+1}) - F_n^*(t_j) - \Delta_j) = (\Delta_j\pi_n - n\Delta_j)/\sqrt{n}$, entonces

$$\mathbf{M}e^{iw^nu^T} = \exp\left\{-i\sum_{j=0}^{m} u_j\sqrt{n}\Delta_j\right\}\mathbf{M}\exp\left\{\frac{i}{\sqrt{n}}\sum_{j=0}^{m} u_j\Delta_j\pi_n\right\} = e^{i\sqrt{n}U}\left(\sum_{j=0}^{m} e^{iu_j/\sqrt{n}}\Delta_j\right)^n.$$

De aquí, para u fijo, utilizando las igualdades

$$e^\alpha = 1 + \alpha + \alpha^2/2 + O(\alpha^3), \quad \ln(1 + \alpha) = \alpha - \alpha^2/2 + O(\alpha^3),$$

cuando $\alpha = O(1)$, hallamos

$$\ln \mathbf{M}e^{iw^nu^T} = -iU\sqrt{n} + n\ln\left[1 - \sum_{j=0}^{m}(1 - e^{iu_j/\sqrt{n}})\Delta_j\right] =$$

$$= -iU\sqrt{n} + n\ln\left[1 + \sum_{j=0}^{m}\left(i\frac{u_j}{\sqrt{n}} - \frac{u_j^2}{2n} + O(n^{-3/2})\right)\Delta_j\right] =$$

$$= -iU\sqrt{n} + n\left[\frac{\sqrt{n}}{iU} - \frac{1}{2n}\sum_{j=0}^{m} u_j^2\Delta_j + \frac{U^2}{2n} + O(n^{-3/2})\right] =$$

$$= \frac{1}{2}\left[-\sum_{j=0}^{m} u_j^2\Delta_j + U^2\right] + O(n^{-1/2}).$$

Comparando con (1), vemos que cuando $n \to \infty$,

$$\mathbf{M}e^{iw^n u^T} \to \mathbf{M}e^{iw^0 u^T}. \tag{2}$$

Sólo queda utilizar el teorema de continuidad para las funciones características de las distribuciones multidimensionales (véase [11], p. 148). ◁

Lema 2. *Para cualquier* $\varepsilon > 0$

$$\lim_{n\to\infty} \sup \mathbf{P}(\omega_\Delta(w^n) > \varepsilon) \to 0 \tag{3}$$

para $\Delta \to 0$, *donde* $\omega_\Delta(y)$ *es el módulo de continuidad de la función* $y \in D(0, 1)$: $\omega_\Delta(y) = \sup\limits_{\substack{0 \leqslant t_1 < t_2 \leqslant 1 \\ |t_1 - t_2| \leqslant \Delta}} |y(t_1) - y(t_2)|$.

Demostración. Sin limitar la generalidad, sólo podemos examinar los números binarios racionales $\Delta = 2^{-l}$. Para $m > 1$ tenemos

$$\omega_\Delta(w^n) \leqslant \omega_\Delta^{[m]} + 2 \max_{k \leqslant 2^m} \omega \left(\frac{k-1}{2^m}, \frac{k}{2^m} \right),$$

donde

$$\omega_\Delta^{[m]} = \max_{\left|\frac{k-j}{2^m}\right| \leqslant \Delta} \left| w^n\left(\frac{k}{2^m}\right) - w^n\left(\frac{j}{2^m}\right) \right|,$$

$$\omega\left(\frac{k-1}{2^m}, \frac{k}{2^m}\right) = \sup_{\frac{k-1}{2^m} \leqslant u < \frac{k}{2^m}} \left| w^n\left(\frac{k}{2^m}\right) - w^n(u) \right|.$$

Para demostrar (3) examinemos

$$\mathbf{P}(\omega_\Delta(w^n) \geqslant 3_\varepsilon) \leqslant \mathbf{P}\omega_\Delta^{[m]} \geqslant \varepsilon) + \mathbf{P}\left(\bigcup_{k=1}^{2^m} \left\{ \omega\left(\frac{k-1}{2^m}, \frac{k}{2^m}\right) \geqslant \varepsilon \right\} \right). \tag{4}$$

Aquí tomemos el primer sumando. Es fácil notar que cuando $l > 3$ el suceso

$$\bigcap_{r=1}^{m} \bigcap_{k=1}^{2^r} \left\{ \left| w^n\left(\frac{k}{2^r}\right) - w^n\left(\frac{k-1}{2^r}\right) \right| < \frac{\varepsilon}{r^2} \right\}$$

provoca $\{\omega_\Delta^{[m]} < \varepsilon\}$. En vista de que para los sucesos adicionales tiene lugar la inclusión inversa, entonces

$$\mathbf{P}(\omega_\Delta^{[m]} \geqslant \varepsilon) \leqslant \mathbf{P}\left(\bigcup_{r=1}^{m} \bigcup_{k=1}^{2^r} \left\{ \left| w^n\left(\frac{k}{2^r}\right) - w^n\left(\frac{k-1}{2^r}\right) \right| \geqslant \frac{\varepsilon}{r^2} \right\} \right). \tag{5}$$

Pero $w^n\left(\frac{k}{2^r}\right) - w^n\left(\frac{k-1}{2^r}\right) = \sqrt{n}\left(F_n^*\left(\frac{k}{2^r}\right) - F_n^*\left(\frac{k-1}{2^r}\right) - \frac{1}{2^r} \right)$, donde $n\left(F_n^*\left(\frac{k}{2^r}\right) - F_n^*\left(\frac{k-1}{2^r}\right) \right)$ es la frecuencia con que los elementos de la muestra van a parar al intervalo cuya longitud constituye 2^{-r}. Con otras palabras, esta es la suma S_n de variables aleatorias en el esquema de Bernoulli con n pruebas y con una probabilidad del caso 1 igual a $p = 2^{-r}$. Como (véase [11], p. 105)

$$\mathbf{M}(S_n - np)^4 = n(p(1-p)^4 + (1-p)p^4) + 3n(n-1)p^2(1-p)^2 \leqslant np + 3n^2p^2,$$

entonces, según la desigualdad del tipo de Chébishev,

$$\mathbf{P}\left(\left|w^n\left(\frac{k}{2^r}\right) - w^n\left(\frac{k-1}{2^r}\right)\right| \geqslant \frac{\varepsilon}{r^2}\right) = \mathbf{P}\left(|S_n - np| \geqslant \frac{\varepsilon\sqrt{n}}{r^2}\right) \leqslant$$

$$\leqslant \frac{(np + 3n^2p^2)r^8}{\varepsilon^4 n^2} = \frac{r^8}{\varepsilon^4 2^r n} + \frac{3r^8}{\varepsilon^4 2^{2r}}.$$

Por consiguiente, el segundo miembro en (5) no supera

$$\sum_{r=1}^{m}\left[\frac{r^8}{\varepsilon^4 n} + \frac{3r^8}{\varepsilon^4 2^r}\right] \leqslant c\left(\frac{m^9}{\varepsilon^4 n} + \frac{l^8}{\varepsilon^4 2^l}\right),$$

donde c es cierta constante absoluta $\left(\sum_{r=1}^{m} r^8 \sim m^9/9\right.$ cuando $m \to \infty$ y $\left.\sum_{r=1}^{\infty} r^8 2^{-r} \sim 2l^8 2^{-l} \text{ cuando } l \to \infty\right)$. Poniendo $m = 3\log_2 n$, obtenemos que

$$\limsup_{n\to\infty} \mathbf{P}(\omega_\Delta^{[m]} \geqslant \varepsilon) \leqslant c\,\frac{l^8}{\varepsilon^4 2^l}.$$

Eligiendo l (o Δ), esta expresión puede hacerse tan pequeña como se quiera.

Ahora apreciemos el segundo sumando en (4), que no supera

$$2^m\mathbf{P}\left(\omega\left(\frac{k-1}{2^m}, \frac{k}{2^m}\right) \geqslant \varepsilon\right). \tag{6}$$

El suceso que aquí figura bajo el signo de probabilidad significa que, elegido m, en el intervalo $((k-1)/n^3,\ k/n^3)$ cuya anchura es n^{-3}, la desviación de $n(F_n^*(u) - u)$ respecto a $n(F_n^*(k/n^3) - k/n^3)$ supera $\sqrt{n}\varepsilon$. En vista de que $\sqrt{n}\varepsilon \geqslant 3$, cuando n es bastante grande, para esto, en el intervalo $((k-1)/n^3,\ k/n^3)$ deben caer por lo menos 2 elementos de la muestra X, o sea, debe producirse el suceso $\{S_n \geqslant 2\}$ si volvemos a utilizar las designaciones para el esquema de Bernoulli cuando $p = n^{-3}$. Pero en vista de que $1 = (1 - p + p)^n = (1 - p)^n + np(1-p)^{n-1} + O(n^2p^2)$, entonces

$$\mathbf{P}(S_n \geqslant 2) = 1 - (1-p)^n - np(1-p)^{n-1} = O(n^2p^2).$$

Ahora bien, (6) no supera $n^3O(n^{-4}) = O(n^{-1}) = O(1)$. El lema queda demostrado. $\lhd$

Demostración del teorema 1. Para cualquier $x \in D(0, 1)$ pongamos

$$\|x\| = \sup_{0\leqslant t\leqslant 1}|x(t)|,\quad f_\varepsilon^+(x) = \sup_{\|y-x\|\leqslant\varepsilon} f(y),\quad f_\varepsilon^-(x) = \inf_{\|y-x\|\leqslant\varepsilon} f(y)$$

y designemos por x_Δ la quebrada continua con nudos en los puntos $(k\Delta,\ x(k\Delta) = x_\Delta(k\Delta))$, $k = 0, \ldots, 1/\Delta$, donde Δ divide por completo 1. Es preciso señalar que

$$\|x - x_\Delta\| \leqslant \omega_\Delta(x) \tag{7}$$

y que $f_\varepsilon^\pm(x_\Delta)$ son funciones continuas del vector $(x(0), x(\Delta), x(2\Delta), \ldots, x(1))$. En virtud del lema 1 y del segundo teorema de continuidad cuando $n \to \infty$,

$$f_\varepsilon^\pm(w_\Delta^n) \Rightarrow f_\varepsilon^\pm(w_\Delta^0). \tag{8}$$

Además, de la continuidad de w^0 y de la funcional f se deduce que

$$\|w_\Delta^0 - w^0\| \leqslant \omega_\Delta(w^0) \xrightarrow[P]{} 0 \quad \text{cuando} \quad \Delta \to 0, \tag{9}$$

$$f_\varepsilon^\pm(w^0) \xrightarrow[P]{} f(w^0) \quad \text{cuando} \quad \varepsilon \to 0. \tag{10}$$

De la definición de f_ε^- se desprende que $f_\varepsilon^-(y) \leqslant f(x)$ en el conjunto $\|y - x\| \leqslant \varepsilon$. Por eso

$$\mathbf{P}(f(w^n) \leqslant t) \leqslant \mathbf{P}(f_\varepsilon^-(w_\Delta^n) \leqslant t, \quad \|w_\Delta^n - w^n\| \leqslant \varepsilon) + \mathbf{P}(\|w_\Delta^n - w^n\| > \varepsilon) \leqslant$$
$$\leqslant \mathbf{P}(f_\varepsilon^-(w_\Delta^n) \leqslant t) + \mathbf{P}(\omega_\Delta(w^n) > \varepsilon).$$

Pasando aquí al límite para $n \to \infty$ y utilizando (8) y (9), obtenemos

$$\limsup_{n\to\infty} \mathbf{P}(f(w^n) \leqslant t) \leqslant \mathbf{P}(f_\varepsilon^-(w_\Delta^0) \leqslant t) + \limsup_{n\to\infty} \mathbf{P}(\omega_\Delta(w^n) > \varepsilon). \tag{11}$$

Análogamente hallamos

$$\mathbf{P}(f_\varepsilon^-(w_\Delta^0) \leqslant t) \leqslant \mathbf{P}(f_{2\varepsilon}^-(w^0) \leqslant t) + \mathbf{P}(\omega_\Delta(w^0) > \varepsilon).$$

Sustituyamos ahora la última expresión en (11) y pasemos al límite cuando $\Delta \to 0$. Entonces, de (9) y del lema (2) obtenemos que

$$\limsup_{n\to\infty} \mathbf{P}(f(w^n) \leqslant t) \leqslant \mathbf{P}(f_{2\varepsilon}^-(w^0) \leqslant t).$$

De aquí y de (10) se deduce que

$$\limsup_{n\to\infty} \mathbf{P}(f(w^n) \leqslant t) \leqslant \mathbf{P}(f(w^0) \leqslant t).$$

Análogamente se establece la desigualdad inversa

$$\liminf_{n\to\infty} \mathbf{P}(f(w^n) < t) \geqslant \mathbf{P}(f(w^0) < t).$$

Las desigualdades obtenidas significan, evidentemente, que $f(w^n) \Rightarrow f(w^0)$. ◁

Examinemos otro teorema límite funcional para los procesos empíricos, el cual se asemeja mucho al teorema 1.

Supongamos que además de la muestra X de volumen n_1 tenemos una muestra Y de volumen n_2 que no depende de la primera y la cual procede de esa misma disitribución uniforme en $[0, 1]$. En las condiciones de este apartado nos será más cómodo designar por $F_X^*(t)$ y $F_Y^*(t)$ las funciones empíricas de distribución de las muestras X e Y, respectivamente. Pongamos

$$w_{X,Y}(t) = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}\,(F_X^*(t) - F_Y^*(t)).$$

Teorema 2. *Si la funcional f satisface las condiciones del teorema* 1, *entonces, para $n_1 \to \infty$, y $n_2 \to \infty$*

$$f(w_{X,Y}) \Rightarrow f(w^0).$$

Demostración. Demostremos este teorema utilizando la suposición simplificadora de que

$$a = \frac{n_2}{n_1 + n_2} \to a_0 \in [0, 1]$$

cuando $n \to \infty$. Tenemos

$$w_{X,Y}(t) = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}\,[(F_X^*(t) - t) - (F_Y^*(t) - t)] = \sqrt{a}\,w_X(t) + \sqrt{1 - a}\,w_Y(t), \tag{12}$$

donde $w_X(t)$ y $w_Y(t)$ son los procesos empíricos que corresponden a las muestras X e Y.

Como $\omega_\Delta(x + y) \leqslant \omega_\Delta(x) + \omega_\Delta(y)$, entonces, de (12) y del lema 2 se deduce inmediatamente el análogo del lema 2 para el proceso $w_{X,Y}(t)$: para cualquier $\varepsilon > 0$,

$$\lim_{n\to\infty} \sup \mathbf{P}(\omega_\Delta(w_{X,Y}) > \varepsilon) \to 0.$$

La convergencia de las distribuciones de dimensión finita $w_{X,Y}$ y w^0 también se desprende de (12). En efecto, designemos por $w_{X,Y}$, w_X, w_Y los vectores construidos a base de los procesos $w_{X,Y}(t)$, $w_X(t)$, $w_Y(t)$, exactamente igual que como fue construido el vector w^n a base del proceso $w^n(t)$. Entonces, utilizando la independencia de X e Y y la demostración del lema 1, obtenemos

$$\mathbf{M}e^{iw_{X,Y}u^T} = \mathbf{M}e^{i\sqrt{a}w_X u^T}\mathbf{M}e^{i\sqrt{1-a}w_Y u^T} \to \exp\left\{\frac{a_0 + (1 - a_0)}{2} \times\right.$$

$$\left.\times\left(\sum_{j=0}^{m} u_j^2\Delta_j - U^2\right)\right\} = \exp\left\{-\frac{1}{2}\left(\sum_{j=0}^{m} u_j^2\Delta_j - U^2\right)\right\} = \mathbf{M}e^{iw^0u^T}.$$

En lo demás, la demostración del teorema 2 no se distingue en nada de la del teorema 1. ◁

Suplemento III

Propiedades de las esperanzas matemáticas condicionales

En el § 2.9 hemos citado las propiedades principales de las e.m.c. Más abajo aducimos las demostraciones de estas propiedades que siguen en el mismo orden que en el § 2.9.

1a. $\mathbf{M}(c\xi/\mathfrak{A}) = c\mathbf{M}(\xi/\mathfrak{A})$.

1b. $\mathbf{M}(\xi_1 + \xi_2/\mathfrak{A}) = \mathbf{M}(\xi_1/\mathfrak{A}) + \mathbf{M}(\xi_2/\mathfrak{A})$.

1c. *Si* $\xi_1 \leqslant \xi_2$ *c.s., entonces* $\mathbf{M}(\xi_1/\mathfrak{A}) \leqslant \mathbf{M}(\xi_2/\mathfrak{A})$ *c.s.*

Para demostrar la propiedad 1a es necesario convencerse, según la definición 2.9.2, de que

1) $c\mathbf{M}(\xi/\mathfrak{A})$ es una función $\mathfrak{A}$-medible.

2) $\mathbf{M}(c\mathbf{M}(\xi/\mathfrak{A}); A) = \mathbf{M}(c\xi; A)$ para cualquier $A \in \mathfrak{A}$.

El cumplimiento de la primera propiedad es evidente. La segunda propiedad se deduce de las propiedades de linielidad de una esperanza matemática ordinaria (o de una integral ordinaria):

$$\mathbf{M}(c\mathbf{M}(\xi/\mathfrak{A}); A) = c\mathbf{M}(\mathbf{M}(\xi/\mathfrak{A}); A) = c\mathbf{M}(\xi; A) = \mathbf{M}(c\xi; A).$$

La propiedad 1b se demuestra exactamente igual.

Para demostrar la propiedad 1c pongamos, para abreviar, $\hat{\xi}_i = \mathbf{M}(\xi_i/\mathfrak{A})$. Entonces, para cualquier $A \in \mathfrak{A}$,

$$\int_A \hat{\xi}_1 d\mathbf{P} = \mathbf{M}(\hat{\xi}_1; A) = \mathbf{M}(\xi_1; A) \leqslant \mathbf{M}(\xi_2; A) = \int_A \hat{\xi}_2 d\mathbf{P},$$

$$\int_A (\hat{\xi}_2 - \hat{\xi}_1) d\mathbf{P} \geqslant 0.$$

De aquí se deduce que $\hat{\xi}_2 - \hat{\xi}_1 \geqslant 0$ c.s.

2. Desigualdad de Chébishev. *Si* $\xi \geqslant 0$, $x > 0$, *entonces*

$$\mathbf{P}(\xi) \geqslant x/\mathfrak{A}) \leqslant \frac{\mathbf{M}(\xi/\mathfrak{A})}{x}.$$

Esta propiedad se desprende de 1c, ya que $\mathbf{P}(\xi \geqslant x/\mathfrak{A}) = \mathbf{M}(I_{\{\xi \geqslant x\}}/\mathfrak{A})$, donde I_A es el indicador del suceso A, y es válida la desigualdad $I_{\{\xi \geqslant x\}} \leqslant \xi/x$.

3. *Si* $\mathfrak{A}$ *y* $\sigma(\xi)$ *son independientes, entonces* $\mathbf{M}(\xi/\mathfrak{A}) = \mathbf{M}\xi$. Como $\hat{\xi} = \mathbf{M}\xi$ es una función $\mathfrak{A}$-medible, sólo nos queda comprobar la segunda condición de definición 2.9.2: para cualquier $A \in \mathfrak{A}$

$$\mathbf{M}(\hat{\xi};\ A) = \mathbf{M}(\xi;\ A).$$

La validez de esta igualdad se deduce de la independencia de las variables aleatorias I_A y ξ y de las relaciones

$$\mathbf{M}(\xi;\ A) = \mathbf{M}(\xi I_A) = \mathbf{M}\xi \cdot \mathbf{M}I_A = \hat{\xi}\mathbf{P}(A) = \mathbf{M}(\hat{\xi};\ A).$$

4. Teorema de convergencia monótona. *Si* $0 \leqslant \xi_n \uparrow \xi$ *c.s., entonces* $\mathbf{M}(\xi_n/\mathfrak{A}) \uparrow \mathbf{M}(\xi/\mathfrak{A})$ *c.s.* En efecto, de $\xi_{n+1} \geqslant \xi_n$ c.s. resulta $\hat{\xi}_{n+1} \geqslant \hat{\xi}_n$ c.s., donde $\hat{\xi}_n = \mathbf{M}(\xi_n/\mathfrak{A})$. Por eso existe una $\mathfrak{A}$-medible $\hat{\xi}$ tal que $\hat{\xi}_n \uparrow \xi$ c.s. En virtud del teorema ordinario de convergencia monótona, para cualquier $A \in \mathfrak{A}$,

$$\int_A \hat{\xi}_n d\mathbf{P} \to \int_A \hat{\xi} d\mathbf{P}, \quad \int_A \xi_n d\mathbf{P} \to \int_A \xi d\mathbf{P}.$$

En vista de que los primeros miembros de estas relaciones coinciden, también coinciden los segundos. Esto precisamente significa que $\hat{\xi} = \mathbf{M}(\xi/\mathfrak{A})$.

5. *Si* η *es real y* $\mathfrak{A}$ *es medible, entonces*

$$\mathbf{M}(\eta\xi/\mathfrak{A}) = \eta\mathbf{M}(\xi/\mathfrak{A}). \tag{1}$$

Si $\eta = I_B$ (indicador del conjunto $B \in \mathfrak{A}$), entonces, la afirmación es justa, ya que para cualquier $A \in \mathfrak{A}$

$$\int_A \mathbf{M}(I_B\xi/\mathfrak{A})d\mathbf{P} = \int_A I_B\xi d\mathbf{P} = \int_{AB} \xi d\mathbf{P} = \int_{AB} \mathbf{M}(\xi/\mathfrak{A})d\mathbf{P} = \int_A I_B\mathbf{M}(\xi/\mathfrak{A})d\mathbf{P}.$$

De aquí y de la linealidad de las e.m.c. resulta que la afirmación también es válida para cualesquiera funciones simples η.

Si $\xi \geqslant 0$ y $\eta \geqslant 0$, entonces, tomando la sucesión de funciones simples $0 \leqslant \eta_n \uparrow \eta$ y haciendo uso del teorema de convergencia monótona en la igualdad

$$\mathbf{M}(\eta_n\xi/\mathfrak{A}) = \eta_n\mathbf{M}(\xi/\mathfrak{A}),$$

obtenemos (1). El paso al caso de ξ y η arbitrarias se realiza ordinariamente: examinando las partes positivas y negativas de las variables aleatorias ξ y η. En este caso, para que las diferencias y sumas obtenidas tengan sentido, es necesario exigir la existencia de $\mathbf{M}|\xi| < \infty$, $\mathbf{M}|\xi_\eta| < \infty$.

6. *La desigualdad de Cauchy — Buniakovski*

$$\mathbf{M}(\xi_1\xi_2/\mathfrak{A}) \geqslant [\mathbf{M}(\xi_1^2/\mathfrak{A})\mathbf{M}(\xi_2^2/\mathfrak{A})]^{1/2}$$

se demuestra exactamente igual que para las esperanzas matemáticas ordinarias (véase, por ejemplo, [11]), puesto que la demostración, además de la linealidad, no utiliza otras propiedades de las esperanzas matemáticas.

La desigualdad de Jensen

$$g(\mathbf{M}(\xi/\mathfrak{A})) \leqslant \mathbf{M}(g(\xi)/\mathfrak{A}) \tag{2}$$

para cualquier función g convexa hacia abajo se deduce de las siguientes relaciones (compárese con [11]). En virtud de la convexidad de $g(x)$, para cada y habrá un número $g_1(y)$ tal, que

$$g(x) \leqslant g(y) + (x - y)g_1(y).$$

Pongamos aquí $x = \xi$, $y = \hat{\xi} = \mathbf{M}(\xi/\mathfrak{A})$ y tomemos la e.m.c. de ambos miembros de esta desigualdad. Como, en virtud de la propiedad 5,

$$\mathbf{M}[(\xi - \hat{\xi})g_1(\hat{\xi})/\mathfrak{A}] = g_1(\hat{\xi})\mathbf{M}[\xi - \hat{\xi}/\mathfrak{A}] = 0,$$

obtenemos (2).

7. *La fórmula de la probabilidad completa* se desprende de la propiedad 8 si en calidad de $\mathfrak{A}$ se adopta la σ-álgebra trivial.

8. *Si* $\mathfrak{A} \subset \mathfrak{A}_1$ $\mathfrak{F}$, entonces es válida la fórmula de "promediación sucesiva"

$$\mathbf{M}(\xi/\mathfrak{A}) = \mathbf{M}(\mathbf{M}(\xi/\mathfrak{A}_1)/\mathfrak{A}).$$

En efecto, para cualquier $A \in \mathfrak{A}$, en virtud de que $A \in \mathfrak{A}_1$,

$$\int_A \mathbf{M}(\mathbf{M}(\xi/\mathfrak{A}_1)/\mathfrak{A})d\mathbf{P} = \int_A \mathbf{M}(\xi/\mathfrak{A}_1)d\mathbf{P} = \int_A \xi d\mathbf{P} = \int_A \mathbf{M}(\xi/\mathfrak{A})d\mathbf{P}.$$

En conclusión cabe señalar que la propiedad 5 admite, para suposiciones amplias, la siguiente generalización.

5A. *Si* η *es medible respecto a* $\mathfrak{A}$, *y* $\varphi(\omega, \eta)$ *es la función medible de las variables* $\omega \in \Omega$ *y* $\eta \in R^k$, *entonces*

$$\mathbf{M}(\varphi(\omega, \eta)/\mathfrak{A}) = \psi(\omega, \eta), \textit{ donde } \psi(\omega, y) = \mathbf{M}(\varphi(\omega, y)/\mathfrak{A}). \tag{3}$$

Demostremos esta propiedad suponiendo que existe una sucesión de funciones simples η_n tal, que $\varphi(\omega, \eta_n)\uparrow\varphi(\omega, \eta)$, $\psi(\omega, \eta_n)\uparrow\varphi(\omega, \eta)$ c.s. En efecto, supongamos que $\eta_n = y_k$ para $\omega \in A_k \subset \mathfrak{A}$. Entonces

$$\varphi(\omega, \eta_n) = \sum_k \varphi(\omega, y_k)I_{A_k}.$$

En virtud de la propiedad 5, de aquí se deduce el cumplimiento de (3) para las funciones η_n. Queda utilizar el teorema de convergencia monótona (propiedad (4) en la igualdad

$$\mathbf{M}(\varphi(\omega, \eta_n)/\mathfrak{A}) = \psi(\omega, \eta_n).$$

Suplemento IV

Teorema de factorización de Neyman — Fisher

En este apartado demostraremos el teorema 2.12.1.

Para simplificar las designaciones supondremos, sin limitar la generalización, que $n = 1$ (pues la muestra X puede ser multidimensional). Además, en concordancia con el acuerdo de que el espacio probabilístico $(\mathscr{X}, \mathfrak{B})$ es muestral, escribiremos $\mathbf{P}_\theta(B)$ en vez de $\mathbf{P}_\theta(X \in B)$ y designaremos por l la dimensión de la estadística S.

Teorema 1. *Supongamos que se cumple la condición* (A_μ). *La estadística S es suficiente si y sólo si existe la función no negativa* $\psi(\theta, s)$ *medible respecto a* $s \in R^l$ *y la función no negati-*

va $h(x)$ medible respecto a $x \in \mathscr{X}$, tales que

$$f_\theta(x) = \frac{d\mathbf{P}_\theta}{d_\mu}(x) = \psi(\theta, S(x))h(x) \quad \text{c.d.} \quad [\mu]. \tag{1}$$

A la demostración del teorema 1 le antepondremos dos afirmaciones auxiliares. Introduzcamos en el planteamiento la

Condición (*D*). *La familia $\mathscr{P} = \{\mathbf{P}_\theta\}_{\theta\in\Theta}$ satisface la condición* (A_λ) (o sea, es dominada por la medida λ), *donde la medida probabilística λ tiene la forma siguiente*:

$$\lambda = \sum_i c_i \mathbf{P}_{\theta_i}, \theta_i \in \Theta, \quad c_i > 0, \quad \sum_i c_i = 1.$$

Teorema 2. *La condición (A_μ) es necesaria y suficiente para el cumplimiento de la condición D.*

Demostración. La necesidad es evidente. Demostremos la suficiencia. Sin limitar la generalidad se puede considerar que μ es una medida probabilística. En efecto, en vez de μ siempre se puede introducir la medida

$$\mu^*(A) = \sum_j \frac{\mu(AB_j)}{2^j_\mu(B_j)},$$

donde $\{B_j\}$ forma la partición del espacio $\mathscr{X}$, tal, que $\mu(B_j) < \infty$, $j = 1, 2, \ldots$

Sea $\tilde{\mathscr{P}}$ la clase de todas las medidas probabilísticas de forma $\mathbf{P} = \sum c_i \mathbf{P}_{\theta_i}$, $\theta_i \in \Theta$, $c_i > 0$, $\sum c_i = 1$. Evidentemente, $\mathscr{P} \subset \tilde{\mathscr{P}}$ y $\tilde{\mathscr{P}}$ también satisface la condición (A_μ).

Designemos $p = d\mathbf{P}/d_\mu$ y examinemos la clase $\mathfrak{S}$ de conjuntos $C \in \mathfrak{B}$ para los cuales existe $\mathbf{P} \in \tilde{\mathscr{P}}$ tal, que $p(x) > 0$ c.d. en C, $\mathbf{P}(C) > 0$. Sea $c_1, C_2, \ldots$ una sucesión de conjuntos de $\mathfrak{S}$ tal, que

$$\mu(C_i) \to \sup_{C\in\mathfrak{S}} \mu(C).$$

Como $C_i \in \mathfrak{S}$, entonces existe $\mathbf{P}^{(i)} \in \tilde{\mathscr{P}}$ tal, que $p^{(i)} = \dfrac{d\mathbf{P}^{(i)}}{d_\mu} > 0$ c.s. en C_i. Pongamos

$$C_0 = \cup C_i, \quad \mathbf{P}^{(0)} = \sum_i c_i \mathbf{P}^{(i)}, \quad p^{(0)} = \sum_i c_i p^{(i)}$$

para cualesquiera $c_i > 0$, $\sum c_i = 1$. Es evidente que $p^{(0)} > 0$ en C_0 y, por lo tanto, $C_0 \in \mathfrak{S}$.

La afirmación del teorema quedará demostrada si determinamos que $\mathbf{P}^{(0)}(A) = 0$ contribuye a que $\mathbf{P}(A) = 0$ para todas $\mathbf{P} \in \tilde{\mathscr{P}}$. Esto significará la continuidad absoluta de $\mathbf{P}_\theta$ respecto a $\lambda = \mathbf{P}^{(0)}$ y el cumplimiento de la condición (*D*).

Así pues, supongamos que $\mathbf{P}^{(0)}A = 0$ y que $\mathbf{P}$ es cualquier otro elemento de $\tilde{\mathscr{P}}$. Designemos $C = \{x: p(x) > 0\}$. La afirmación requerida se deducirá de las tres relaciones siguientes:

$$\mathbf{P}(AC_0) = 0, \quad \mathbf{P}(A\overline{C_0}\overline{C}) = 0, \quad \mathbf{P}(A\overline{C_0}) = 0,$$

donde $\overline{B}$ significa el complemento de B. La primera de estas relaciones se desprende del hecho de que $\mathbf{P}^{(0)}(AC_0) = 0$, $p^{(0)}(x) > 0$ en C_0 y, por lo tanto, $\mu(AC_0) = 0$. La segunda relación resulta del hecho de que $p(x) = 0$ en $\overline{C}$. Para demostrar la tercera relación admitamos que ella es injusta. Entonces, poniendo $R = A\overline{C_0}C$, obtenemos $\mu(R) > 0$, $\mu(C_0 \cup R) - \mu(C_0 > 0$. Pero esto contradice la igualdad

$$\mu(C_0) = \sup_{C\in\mathfrak{C}} \mu(C),$$

en vista de que $C_0 \in \mathfrak{C}$, $R \in \mathfrak{C}$, $C_0 \cup R \in \mathfrak{C}$. $\lhd$

Ahora bien, hemos establecido que, al cumplirse las condiciones (A_μ), existe una medida λ para la cual se cumple la condición (D).

Teorema 3. *La estadística S es suficiente si y sólo si existe una función medible* $g_\theta(s)$ *tal, que*

$$\frac{d\mathbf{P}_\theta}{d\lambda}(x) = g_\theta(S(x)) \text{ c.d. } [\lambda]. \tag{2}$$

Demostración. Para cualquier $B \subset R^l$ medible designemos $S^{-1}(B) = \{x \in \mathscr{X}: S(x) \in B\} \in \in \mathfrak{B}_{\mathscr{X}}$ y examinemos la distribución G_θ en R^l de la estadística S, inducida por la distribución $\mathbf{P}_\theta$:

$$G_\theta(B) = \int_{S^{-1}(B)} \mathbf{P}_\theta(dx) = \int_{S^{-1}(B)} \frac{d\mathbf{P}_\theta}{d\lambda}(x)\lambda(dx).$$

Examinemos también la distribución

$$\nu(B) = \int_{S^{-1}(B)} \lambda(dx).$$

Por supuesto que G_θ es absolutamente continua respecto a ν, ya que $\nu(B) = 0$ contribuye a que $G_\theta(B) = 0$. Por eso existe una densidad $g_\theta(s)$ medible en s, tal, que

$$G_\theta(B) = \int_B g_\theta(s)\nu(ds).$$

Ahora supongamos que S es una estadística suficiente y, por consiguiente, que existe una variante de distribución condicional $\mathbf{P}(A/s) = \mathbf{P}_\theta(A/S(x) = s)$ que no depende de θ. Según la definición de la distribución convencional, para cualquier $A_0 \in \sigma(S)$ se cumple

$$\int_{A_0} \mathbf{P}(A/S(x))\mathbf{P}_\theta(dx) = \mathbf{P}_\theta(A \cap A_0).$$

De aquí también se deduce que

$$\int_{A_0} \mathbf{P}(A/S)(x))\lambda(dx) = \lambda(A \cap A_0).$$

Esto significa que $\mathbf{P}(A/S)$ es a la vez una probabilidad condicional respecto a la distribución λ. Designemos esta probabilidad como e.m.c. $E_\lambda(I_A/S)$ del indicador I_A.

De (1), cuando $A_0 = R^l$, en virtud de las propiedades de la e.m.c., obtenemos

$$\mathbf{P}_\theta(A) = \int \mathbf{P}(A/S(x))\mathbf{P}_\theta(dx) = \mathbf{M}_\theta\mathbf{P}(A/S(X)) =$$
$$= \int \mathbf{P}(A/S)G_\theta(ds) = \int \mathbf{P}(A/S)g_\theta(s)\nu(ds) = \int \mathbf{P}(A/S(x))g_\theta(S(x))\lambda(dx) =$$
$$= \int \mathbf{E}_\lambda(I_A/S(x))g_\theta(S(x))\lambda(dx) = \int \mathbf{E}_\lambda(I_A g_\theta(S(x))/S(x))\lambda(dx) =$$
$$= \int I_A g_\theta(S(x))\lambda(dx) = \int_A g_\theta(S(x))\lambda(dx).$$

Es evidente que esto significa precisamente (2).

Ahora supongamos que se cumple (2). Demostremos que la e.m.c. $\mathbf{E}_\lambda(I_A/S)$, correspondiente a la distribución λ (que no depende de θ), es a la vez la e.m.c. $\mathbf{P}_\theta(A/S)$ para todas $\mathbf{P}_\theta \in \mathscr{P}$.

Para A y θ fijos introduzcamos la medida γ en $\mathfrak{B}$, definiéndola por la igualdad

$$\gamma(C) = \mathbf{P}_\theta(AC), \quad C \in \mathfrak{B},$$

así que $d\gamma/d\mathbf{P}_\theta = I_A$, $d\gamma/d\lambda = I_A g_\theta(S(x))$.

Para cualquier $C \in \sigma(S)$ tenemos

$$\gamma(C) = \int_C I_A \mathbf{P}_\theta(dx) = \mathbf{M}_\theta I_A I_C = \mathbf{M}_\theta I_C \mathbf{M}_\theta(I_A/S) = \int_C \mathbf{M}_\theta(I_A/S)\mathbf{P}_\theta(dx). \tag{3}$$

Por consiguiente, si γ, $\mathbf{P}_\theta$, λ se examinan como distribuciones en $\sigma(S)$, entonces

$$\frac{d\gamma}{d\mathbf{P}_\theta} = \mathbf{M}_\theta(I_A/S),$$

$$\frac{d\gamma}{d\lambda} = \mathbf{M}_\theta(I_A/S)\frac{d\mathbf{P}_\theta}{d\lambda} = \mathbf{M}_\theta(I_A/S)g_\theta(S).$$

Análogamente, en virtud de (3), en $\sigma(S)$,

$$\frac{d\gamma}{d\lambda} = \mathbf{E}_\lambda(I_A g_\theta(S)/S) = g_\theta(S)\mathbf{E}_\lambda(I_A/S).$$

De aquí se deduce que λ casi seguramente (aquí y más adelante, por λ y $\mathbf{P}_\theta$ entenderemos las distribuciones en $\sigma(S)$) constituirá

$$\mathbf{M}_\theta(I_A/S)g_\theta(S) = \mathbf{E}_\lambda(I_A/S)g_\theta(S). \tag{4}$$

Ahora hagamos uso de la propiedad (D), en virtud de la cual el cumplimiento de (4) λ c.s. significa el cumplimiento de esta relación por $\mathbf{P}_\theta$ c.s. Además, $\mathbf{P}_\theta$ c.s. es

$$g_\theta(S(x)) = \frac{d\mathbf{P}_\theta}{d\lambda}(x) \neq 0.$$

Por consiguiente, $\mathbf{P}_\theta$ c.s. es válida,

$$\mathbf{P}_\theta(A/S) = \mathbf{M}_\theta(I_A/S) = \mathbf{E}_\lambda(I_A/S).$$

Esto significa que la magnitud $\mathbf{E}_\lambda(I_A/S)$, que no depende de θ, puede ser elegida en calidad de probabilidad condicional $\mathbf{P}_\theta(A/S)$. $\lhd$

Demostración del teorema 1. Si S es una estadística suficiente, entonces (1) se deduce del teorema 3, ya que

$$f_\theta(x) = \frac{d\mathbf{P}_\theta}{d\mu} = g_\theta(S(x))\frac{d\lambda}{d\mu}(x),$$

donde es preciso suponer que $g_\theta(s) = \psi(\theta, s)$, $\dfrac{d\lambda}{d\mu}(x) = h(x)$. Al contrario, si (1) es válida, entonces

$$\frac{d\lambda}{d\mu} = \sum c_i \frac{d\mathbf{P}_{\theta_i}}{d\mu} = \sum c_i \psi(\theta_i, S(x))h(x) = r(S(x))h(x).$$

Por eso, si $r(S(x)) > 0$, entonces

$$\frac{d\mathbf{P}_\theta}{d\lambda} = \frac{d\mathbf{P}_\theta}{d\mu}\cdot\frac{d\mu}{d\lambda} = \frac{\psi(\theta, S(x))}{r(S(x))}.$$

Si $r(S(x)) = 0$, entonces, $\dfrac{d\mathbf{P}}{d\lambda}(x)$ se puede definir arbitrariamente, ya que λ-medida y, por consiguiente, $\mathbf{P}_\theta$-medida del conjunto de tales puntos x es igual a cero. Poniendo $g_\theta(s) = \psi(\theta, s)/r(s)$ y aplicando el teorema 3 obtenemos que S es una estadística suficiente. $\lhd$

Suplemento V

Ley de los grandes números y teorema central del límite. Variantes uniformes

1. Ley de los grandes números en el esquema de series. Examinemos las sucesiones $\{\xi_{k,n}\}_{k=1}^{n}$, $n = 1, 2, \ldots$, de vectores igualmente distribuidos en el esquema de series (la distribución $\xi_{k,n}$ depende de n) y supongamos que $\mathbf{M}\xi_{k,n} = 0$.

Designemos $\zeta_n = \sum_{k=1}^{n} \xi_{k,n}$.

Teorema 1. *Sea*

$$n\mathbf{M}|\xi_{k,n}| = a_n < a < \infty,$$
$$n\mathbf{M}(|\xi_{k,n}|;\ |\xi_{k,n}| > \tau) \to 0 \tag{1}$$

cuando $n \to \infty$ para cualquier $\tau > 0$. Entonces, para cualquier $\varepsilon > 0$,

$$\mathbf{P}(|\zeta_n| > \varepsilon) \to 0.$$

Demostración. Examinemos los cortes $\xi'_{k,n}$ de las variables aleatorias $\xi_{k,n}$ en el nivel τ:

$$\xi'_{k,n} = \begin{cases} \xi_{k,n}, & \text{si} \quad |\xi_{k,n}| \leqslant \tau, \\ 0, & \text{si} \quad |\xi_{k,n}| > \tau. \end{cases}$$

En virtud de la condición (1)

$$\mathbf{P}(\xi'_{1,n} \neq \xi_{1,n}) = \mathbf{P}(|\xi_{1,n}| > \tau) \leqslant \frac{1}{\tau}\mathbf{M}(|\xi_{1,n}|;\ |\xi_{1,n}| > \tau) = O(1/n),\ \mathbf{M}\xi'_{1,n} = O(1/n),$$

$$\mathbf{M}(\xi'_{1,n})^2 = \mathbf{M}(\xi^2_{1n};\ |\xi_{1,n}| \leqslant \tau) \leqslant$$
$$\leqslant \tau\mathbf{M}(|\xi_{1,n}|;\ |\xi_{1,n}| \leqslant \tau) = \tau(a_n/n - \mathbf{M}(|\xi_{1,n}|;\ |\xi_{1,n}| > \tau)).$$

Por eso, para cualquier $\varepsilon > 0$ y para valores bastante grandes de n,

$$\mathbf{M}(\xi'_{1,n})^2 \leqslant 2a\tau/n,\ \mathbf{D}\xi'_{1,n} \leqslant 2a\tau/n,\ n\mathbf{M}\xi'_{1,n} < \varepsilon/2.$$

Pongamos $\zeta'_n = \sum_{j=1}^{n} \xi'_{j,n}$. Entonces, si los valores de n son bastante grandes,

$$\mathbf{P}(|\zeta_n| > \varepsilon) \leqslant \mathbf{P}\left(\bigcup_{j=1}^{n} \{\xi'_{j,n} \neq \xi_{j,n}\}\right) + \mathbf{P}(|\zeta'_n| > \varepsilon).$$

Aquí, el primer sumando no supera $n\mathbf{P}(\xi'_{1,n} \neq \xi_{1,n}) = o(1)$, y el segundo no pasa de

$$\mathbf{P}(|\zeta'_n - \mathbf{M}\zeta'_n| > \varepsilon/2) \leqslant 4\mathbf{D}\zeta'_n/\varepsilon^2 \leqslant 8a\tau/\varepsilon^2.$$

Como τ es arbitrario, para cualquier $\varepsilon > 0$ dado, el valor obtenido puede hacerse tan pequeño como se quiera. Eligiendo ahora un valor de n bastante grande, también podemos hacer tan pequeña como se quiera toda la probabilidad $\mathbf{P}(|\zeta_n| > \varepsilon)$. ◃

2. Teorema central del límite en el esquema de series. Aquí supondremos que

$$\mathbf{M}\xi_{j,n} = 0,\ \mathbf{M}|\xi_{j,n}|^2 < \infty.$$

Designemos $\sigma_n^2 = n\mathbf{M}\xi^T_{1,n}\xi_{1,n}$, $\zeta_n = \sum_{j=1}^{n} \xi_{j,n}$.

Teorema 2. *Supongamos que se cumplen las condiciones de Lindeberg*

$$n\mathbf{M}(|\xi_{1,n}|^2;\ |\xi_{1,n}| > \tau) \to 0$$

para $n \to \infty$ *para cualquier* $\tau > 0$. *Entonces, si* $\sigma_n^2 \to \sigma^2$,

$$\zeta_n \in \Phi_{0,\sigma^2}.$$

Corolario 1 (teorema central ordinario del límite). *Si* $\xi_1, \xi_2, \ldots$ *es una sucesión de vectores independientes igualmente distribuidos,* $\mathbf{M}\xi_k = 0$, $\sigma^2 = \mathbf{M}\xi_k^T\xi_k < \infty$, $s_n = \sum_{k=1}^{n} \xi_k$, *entonces, para* $n \to \infty$,

$$\frac{s_n}{\sqrt{n}} \in \Phi_{0,\sigma^2}.$$

Esta afirmación es el corolario del teorema 2, ya que las variables aleatorias $\xi_{k,n} = \xi_k/\sqrt{n}$ satisfacen las condiciones del mismo.

Demostración del teorema 2. Examinemos las funciones características

$$\psi_n(t) = \mathbf{M}e^{i(t,\xi_{1,n})}, \quad \varphi_n(t) = \mathbf{M}e^{i(t,\zeta_n)} = \psi_n^n(t).$$

Para demostrar el teorema necesitamos convencernos que para cualquier t

$$\varphi_n(t) \to \exp\left\{-\frac{1}{2}\, t\sigma^2 t^T\right\}$$

cuando $n \to \infty$.

Hagamos uso de la variante unidimensional del teorema 1, demostrada en [11]. Las funciones $\psi_n(t)$ y $\varphi_n(t)$ pueden considerarse como funciones características

$$\psi_n^{(\omega)}(v) = \mathbf{M}e^{iv\xi_{1,n}^{\omega}} \quad \text{y} \quad \varphi_n^{(\omega)}(v) = \mathbf{M}e^{iv\zeta_n^{\omega}}$$

de las variables aleatorias $\xi_{1,n}^{\omega} = (\xi_{1,n}, \omega)$, $\zeta_n^{\omega} = (\zeta_n, \omega)$, donde $\omega = t/|t|$, $v = |t|$.

Mostremos que las variables aleatorias escalares $\xi_{k,n}^{\omega}$ satisfacen las condiciones del teorema 1 para el caso unidimensional. Es evidente que

$$\mathbf{M}\xi_{k,n}^{\omega} = 0,\ n\mathbf{M}(\xi_{1,n}^{\omega})^2 = n\mathbf{M}(\xi_{1,n}, \omega)^2 = \omega\sigma_n^2\omega^T \to \omega_{\sigma^2}\omega^T.$$

El cumplimiento de la condición de Lindeberg se deduce de la desigualdad evidente

$$n\mathbf{M}((\xi_{1,n}, \omega)^2;\ |(\xi_{1,n}, \omega)| > \tau) \leqslant n\mathbf{M}(|\xi_{1,n}|^2;\ |\xi_{1,n}| > \tau).$$

Ahora bien, para cualesquiera v y ω (o sea, para cualesquiera t)

$$\varphi_n(t)\mathbf{M}e^{iv\zeta_n^{\omega}} \to \exp\left\{-\frac{1}{2}\, v^2\omega\sigma^2\omega^T\right\} = \exp\left\{-\frac{1}{2}\, t\sigma^2 t^T\right\}. \ \triangleleft$$

3. Teoremas uniformes del límite para las sumas de las variables aleatorias que dependen del parámetro. En este apartado demostraremos los teoremas 29.1 y 29.2.

Sea $X \in \mathbf{P}_\theta$ y $a(x, \theta)$ una función medible $\mathscr{X} \times \Theta \to R^l$ dada,

$$s_n(\theta) = \sum_{j=1}^{n} a(x_j, \theta).$$

Diremos que la integral $a(\theta) = \int a(x, \theta)\mathbf{P}_\theta(dx)$ converge uniformemente en θ en la región $\Theta_0 \subset \Theta$ si

$$\sup_{\theta\in\Theta_0} \int_{|a(x,\theta)| > N} |a(x, \theta)|\mathbf{P}_\theta(dx) \to 0$$

cuando $N \to \infty$.

Teorema 3. (ley uniforme de los grandes números). *Si la integral* $a(\theta) = \int a(x, \theta)\mathbf{P}_\theta(dx)$ *converge uniformemente en* θ *en la región* $\Theta_0 \subset \Theta$, *entonces*

$$\zeta_n(\theta) = \frac{s_n(\theta)}{n} - a(\theta) \underset{\mathbf{P}}{\rightarrow} 0 \tag{2}$$

uniformementne respecto a $\theta \in \Theta_0$.

Demostración. Supongamos que (2) no tiene lugar. Entonces habrá $\varepsilon > 0$, $\delta > 0$ y una sucesión $\theta_n \in \Theta_0$ tales, que

$$\mathbf{P}_{\theta_n}\left(\left|\frac{\zeta_n(\theta_n)}{n}\right| > \varepsilon\right) > \delta \tag{3}$$

para todos n.

Examinemos las variables aleatorias

$$\xi_{j,n} = \frac{n(x_j, \theta_n) - a(\theta_n)}{n}.$$

No es difícil notar que éstas satisfacen las condiciones del teorema 1. En efecto, pongamos $A_n = \{x;\ |a(x, \theta_n) - a(\theta_n)| > \tau n\}$. Entonces

$$n\mathbf{M}_{\theta_n}|\xi_{j,n}| \leqslant 2a = 2\sup_{\theta\in\Theta_0}\int|a(x, \theta)|\mathbf{P}_\theta(dx) < \infty,$$

$$n\mathbf{M}_{\theta_n}(|\xi_{j,n}|;\ |\xi_{j,n}| > \tau) = \int_{A_n}|a(x, \theta_n) - a(\theta_n)|\mathbf{P}_{\theta_n}(dx) \rightarrow 0.$$

La última relación se deduce de la convergencia uniforme de la integral $a(\theta)$ y de la desigualdad de Chébishev

$$\mathbf{P}_{\theta_n}(A_n) \leqslant \frac{\mathbf{M}_{\theta_n}|\xi_{1,n}|}{\tau} \leqslant \frac{2a}{\tau n} \rightarrow 0.$$

Lo dicho significa que la sucesión $\xi_{j,n}$ satisface la ley de los grandes números:

$$\mathbf{P}_{\theta_n}\left(\left|\sum_{j=1}^{n}\xi_{j,n}\right| > \varepsilon\right) \rightarrow 0$$

para cualquier $\varepsilon > 0$. Esto contradice (3) y demuestra el teorema. ◁

Pasemos al *teorema central del límite.* Sea $\mathbf{M}_\theta a(x_1, \theta) = 0$.

Pongamos $\sigma^2(\theta) = \|\sigma_{ij}(\theta)\| = \mathbf{M}_\theta a^T(x_1, \theta)a(x_1, \theta)$ y designemos por $a_j(x, \theta)$, $j = 1, \ldots$ $\ldots, l$ las coordenadas de los vectores $a(x, \theta)$.

Teorema 4 (Teorema central uniforme del límite). *Supongamos que las integrales* $\sigma_{jj}(\theta) = \mathbf{M}_\theta a_j^2(x_1, \theta)$ *convergen uniformemente en* $\Theta_0 \subset \Theta$, *o sea,*

$$\sup_{\Theta_0}\sigma_{jj}(\theta) < \infty$$

$$\sup_{\Theta_0}\mathbf{M}_\theta(a_j^2(x_1, \theta);\ |a_j(x_1, \theta)| > N) \rightarrow 0$$

cuando $N \rightarrow \infty$. *Entonces*

$$\frac{s_n(\theta)}{\sqrt{n}} \Subset \Phi_{0,\sigma^2(\theta)} \tag{5}$$

cuando $n \rightarrow \infty$ *uniformemente respecto a* $\theta \in \Theta_0$.

Demostración. El incumplimiento de (5) significará la existencia de una sucesión $\theta_n \in \Theta_0$ para la cual las sumas de las variables aleatorias $\xi_{j,n} = a(x_j, \theta_n)/\sqrt{n}$ no se aproximarán, según la distribución, a $\Phi_{0,\sigma^2(\theta_n)}$.

En virtud de la compactibilidad de la clausura $\{\sigma^2(\theta), \theta \in \Theta_0\}$, la sucesión θ_n puede considerarse elegida de tal modo que, para cierta matriz σ^2,

$$\sigma^2(\theta_n) = n\mathbf{M}_{\theta_n}\xi_{1,n}^T\xi_{1,n} \to \sigma^2. \tag{6}$$

Entonces, nuestra suposición acerca del incumplimiento de (5) significará que $\sum_{j=1}^{n} \xi_{j,n}$ no se aproximará, según la distribución, a Φ_{0,σ^2}. Pero esto es imposible en virtud del teorema 2, ya que $\xi_{j,n}$ satisfacen las condiciones del referido teorema. En efecto, en virtud de (6) es suficiente verificar la condición de Lindeberg. Para los conjuntos $A_{l,n} = \{|a_l(x_1, \theta_n)| > \tau\sqrt{n}/l\}$ hallamos

$$\sup_{\theta\in\Theta_0} \mathbf{P}_\theta(A_{l,n}) \leqslant \sup_{\theta\in\Theta_0} \frac{l\sigma_{ll}(\theta)}{n\tau^2} \to 0$$

cuando $n \to \infty$. Utilizando el hecho de que $\{|\xi_{1,n}| > \tau\} \subset \bigcup_{l=1}^{l} A_{l,n}$, obtenemos

$$n\mathbf{M}_{\theta_n}(|\xi_{1,n}|^2;\ |\xi_{1,n}| > \tau) \leqslant \sum_{i,k=1}^{l} \mathbf{M}_{\theta_n}(a_i^2(x_1, \theta_n);\ A_{k,n}). \tag{7}$$

Aquí $\mathbf{M}_{\theta_n}(a_l^2(x_1, \theta_n); A_{l,n}) \to 0$ en virtud de la convergencia uniforme de la integral $\sigma_{ll}(\theta)$. Si $i \neq k$, entonces, poniendo $B_{l,N} = \{|a_l(x_1, \theta_n)| > N\}$, obtenemos

$$\mathbf{M}_{\theta_n}(a_l^2;\ A_{k,n}) = \mathbf{M}_{\theta_n}(a_l^2;\ A_{k,n}B_{l,N}) + \mathbf{M}_{\theta_n}(a_l^2;\ A_{k,n}\overline{B}_{l,N}).$$

Aquí, para $\varepsilon > 0$ dado se puede escoger N de tal modo que el primer sumando, en virtud de (4), sea menor que ε. El segundo sumando no supera $N^2\mathbf{P}_{\theta_n}(A_{k,n}) \to 0$ cuando $n \to \infty$. Esto significa que (7) converge a cero cuando $n \to \infty$.

Suplemento VI

Algunas afirmaciones referentes a las integrales que dependen del parámetro

1. Teoremas de la convergencia de las integrales que dependen del parámetro. Sea $\{\psi(t, y)\}$ una familia de funciones medibles que se dan en el espacio medible $(\mathscr{Y}, \mathfrak{B}_{\mathscr{Y}})$ con la medida ν en el. Nos interesarán las condiciones en las que

$$\int\psi(t, y)\nu(dy) \to \int\psi(\theta, y)\nu(dy) \quad \text{cuando} \quad t \to \theta. \tag{1}$$

Sea $\{A(t) = A(t, \theta), t \in \Theta\}$ cierta familia de conjuntos $\mathfrak{B}_{\mathscr{Y}}$. Designemos por $I_{A(t)}(x)$ el indicador $A(t)$, y por $\overline{A}(t)$, el complemento para $A(t)$.

La siguiente afirmación es cierta generalización del teorema conocido de Lebesgue.

Teorema 1. *Supongamos que la familia $\{A(t)\}$ es tal, que*

1) $\psi(t, y)I_{A(t)}(y) \to \psi(\theta)$, *cuando* $t \to \theta$ *para c.t.*$[\nu]$ *valores de* y, *para los cuales* $\psi(\theta, y) \neq 0$.

2) $\sup_t |\psi(t, y) I_{A(t)}(y) \leq \psi(y)$, *donde* ψ *es la función integrable*

$$\int \psi(y) \nu(dy) < \infty.$$

Entonces, para que se cumpla (1) es necesario y suficiente que

$$\int \psi(t, y) I_{\overline{A(t)}}(y) \nu(dy) \to 0 \quad \text{cuando} \quad t \to 0. \tag{2}$$

Demostración. En virtud del teorema de Lebesgue,

$$\int \psi(t, y) I_{A(t)}(y) \nu(dy) \to \int \psi(\theta, y) \nu(dy).$$

En vista de que

$$\int \psi = \int \psi I_A + \int \psi I_{\overline{A}},$$

(1) es equivalente a (2). ◁

Si existe $\int \psi(\theta, y) \nu(dy)$, entonces, en calidad de conjunto $A(t)$ para c.t. $[\nu]$ de $\psi(t, y)$ continuas, se pueden utilizar los conjuntos

$$A(t) = \{y: |\psi(t, y)| \leq 2 |\psi(\theta, y)|\},$$

así como se hace, por ejemplo, en la afirmación siguiente.

Corolario 1. *Sea* $\pi(x)$ *cualquier función medible limitada* $\mathscr{X}^n \to R$, $f_\theta(x)$, *continua en* θ *para c.t.* $[\mu^n]$ *valores de* $x \in \mathscr{X}^n$. *Entonces, la función*

$$\mathbf{M}_\theta \pi(X) = \int \pi(x) f_\theta(x) \mu^n(dx)$$

será continua en θ.

Demostración. Utilicemos el teorema 1 para $\mathscr{Y} = \mathscr{X}^n$, $y = x$, $\nu = \mu^n$, $\psi(t, x) = \pi(x) f_t(x)$, $A(t) = \{x: f_t(x) \leq 2 f_\theta(x)\}$. Es evidente que se han cumplido las condiciones 1) y 2). Como para $\pi(x) \equiv 1$, la función $\mathbf{M}_\theta \pi(X) \equiv 1$ es continua, entonces se cumple (véase (2))

$$\int_{x \notin A(t)} f_t(x) \mu^n(dx) \to 0$$

cuando $t \to \theta$. Pero de aquí, según el teorema 1, resulta la continuidad de $\mathbf{M}_\theta \pi(X)$ para cualquier función limitada π. ◁

Si sólo se trata de la condición suficiente para la convergencia (1) en caso de $\psi(t, y) \to \psi(\theta, y)$ c.d. cuando $t \to \theta$, en calidad de tal condición se puede utilizar la convergencia uniforme de las integrales en (1). Esta última puede ser definida como la existencia de una medida finita λ tal, que la desigualdad $\lambda(A) < \delta = \delta(\varepsilon)$ contribuye a que $\sup_t \int_A |\psi(t, y)| \nu(dy) < \varepsilon$ para $\varepsilon > 0$ dado.

Si existe la mayorante integrable $\psi(y) = \sup_t \psi(t, y)$, entonces siempre existe tal medida λ: es suficiente suponer que $\lambda(A) = \int_A \psi(y) \nu(dy)$.

2. Corolarios de las condiciones (*R*). Aquí demostraremos el lema 2.16.1 y la convergencia uniforme de la integral $I(\theta)$:

$$\sup_\theta \mathbf{M}_\theta(|l'(x_1, \theta)|^2; \quad |l'(x_1, \theta)| > N) \to 0 \tag{3}$$

cuando $N \to \infty$ (precisamente tal uniformidad se tiene en cuenta en los §§ 2.24, 2.28 y 2.29). En vista de que los planteamientos referentes al parámetro unidimensional y multidimensional prácticamente no se distinguen, en este apartado y en el que le sigue nos limitaremos a estudiar el caso unidimensional.

Teorema 2. (lema 2.16.1). *Supongamos que se cumple la condición (R) y que $S = S(X)$ es cualquier estadística para la cual $\mathbf{M}_\theta S^2 < c < \infty$ cuando $\theta \in \Theta$. Entonces, en la igualdad*

$$a_S(\theta) = \mathbf{M}_\theta(S) = \int S(x) f_\theta(x)\mu^n(dx)$$

es posible la derivación bajo el signo integral:

$$a_S'(\theta) = \int S(x) f_\theta'(x)\mu^n(dx) = \mathbf{M}_\theta SL'(X, \theta), \tag{4}$$

siendo, en este caso, continua la función $a_S'(\theta)$.

Demostración. Nótese previamente que de (4), cuando $S(x) \equiv 1$ y $n = 1$, resulta

$$\int f_\theta'(x)\mu(dx) = 0. \tag{5}$$

Como $L'(X, \theta) = \sum_{i=1}^{n} l'(x_i, \theta)$ es la suma de las variables aleatorias independientes con media nula (véase (5)), entonces

$$\mathbf{D}_\theta L'(X, \theta) = \mathbf{M}_\theta(L'(X, \theta))^2 = n\mathbf{M}_\theta(l'(x_1, \theta))^2 = nI(\theta). \tag{6}$$

Ahora supongamos que la función

$$I_n(\theta) = \mathbf{M}_\theta(L'(X, \theta))^2 = 4\int(\sqrt{f_\theta(x)}\,')^2\mu^n(dx)$$

es continua en θ (aún no podemos utilizar (6)). Hagamos uso ahora del teorema 1 para $\mathscr{Y} = \mathscr{X}^n$, $\nu = \mu^n$, $\psi(t, x) = (\sqrt{f_t(x)}\,')^2$, $\delta = t - \theta$, $A(t) = A_1(\delta) \equiv \{x\colon \sup_{v:|\theta - v| < |\delta|} \sqrt{f_v(x)} <$

$$< 2\sqrt{f_\theta(x)},\ \sup_{v:\ |\theta - v| < |\delta|} |\sqrt{f_v(x)}\,'| \leqslant 2|\sqrt{f_\theta(x)}\,'|\}. \tag{7}$$

Las condiciones 1) y 2) del teorema 1 para $\psi(x) = 2\psi(\theta, x)$ se cumplen en viritud de la continuidad de las funciones $\sqrt{f_\theta}$ y $\sqrt{f_\theta'}$. Por eso, de la convergencia de $I_n(t)$ hacia $I_n(\theta)$ cuando $t \to \theta$ obtenemos (véase (2)) que, cuando $t \to \theta$,

$$\varepsilon(t) = \int\limits_{x \notin A_1(\delta)} (\sqrt{f_t(x)}\,')^2\mu^n(dx) \to 0. \tag{8}$$

Al igual que como hemos obrado en el corolario 1, de aquí obtenemos la continuidad de $\int S(x) f_\theta'(x)\mu^n(dx)$. Para convencernos de ello es necesario valerse del teorema 1 "en sentido inverso" y utilizar los mismos conjuntos $A(t)$ y $\psi(t, x) = S(x) f_t'(x)$. Las condiciones 1) y 2) del teorema 1 serán, evidentemente, cumplidas ($\psi(x) = 2|S(x) f_\theta'(x)|$, $\int \psi(x)\mu^n(dx) \leqslant$ $\leqslant 4\mathbf{M}_\theta S^2 \times \int(\sqrt{f_\theta(x)}\,')^2\mu^n(dx)$). El cumplimiento de (2) es asegurado por (8) y por la desigualdad recién citada, en la que la integración ha de efectuarse con arreglo al conjunto $x \notin A_1(\delta)$.

Ahora recurriremos directamente a la demostración de (4). Nótese que

$$\frac{1}{\delta}\left(\int S f_{\theta+\delta}\mu^n - \int S f_\theta \mu^n\right) = \int\int_0^1 S f_\theta' + u\delta^{du}\mu^n = \int\int_0^1 2S\sqrt{f_{\theta+u\delta}}(\sqrt{f_{\theta+u\delta}})'\,du\mu^n.$$

Utilicemos de nuevo el teorema 1 para $\mathscr{Y} = R \times \mathscr{X}^n$, $y = (u, x)$, $\nu = \lambda \times \mu^n$ (λ es la medida de Lebesgue), $\psi(\delta, y) = S(x) f_{\theta+u\delta}'(x)$, $\delta \to 0$, $A(\delta) = A_1(\delta)$, donde $A_1(\delta)$ ha sido definida en (7). Otra vez de la continuidad de $\sqrt{f_\theta(x)}$ y $\sqrt{f_\theta(x)}\,'$ se deduce el cumplimiento de las condiciones 1) y 2) del teorema 1:

$$\psi(\delta, y) I_{A(\delta)}(x) \to S(x) f_\theta'(x) = \psi(0, y) \quad \text{cuando} \quad \delta \to 0,$$
$$\sup_\delta |\psi(\delta, y) I_{A(\delta)}(x)| \leqslant 4S(x)|f_\theta'(x)|,$$

donde, en virtud de la desigualdad de Cauchy — Buniakovski,

$$\int 4S|f_\theta'|\mu^n \leqslant 4\left[\int S^2 f_\theta \mu^n \cdot \int \frac{(f_\theta')^2}{f_\theta}\mu^n\right]^{1/2} < \infty.$$

Ahora bien, para demostrar (4) necesitaremos verificar la condición (2). Esta se desprende de la desigualdad de Cauchy — Buniakovski y de la relación (8):

$$\int\limits_{x\notin A_1(\delta)} \int\limits_0^1 S\sqrt{f_{\theta+u\delta}}(\sqrt{f_{\theta+u\delta}})'\,du\mu^n| \leqslant$$

$$\leqslant \left[\int \int\limits_0^1 S^2 f_{\theta+u\delta}du\mu^n\right]^{1/2}\left[\int\limits_{x\notin A_1(\delta)} \int\limits_0^1 (\sqrt{f_{\theta+u\delta}})'^2 du\mu^n\right]^{1/2} \leqslant$$

$$\leqslant c^{1/2}\left[\int\limits_0^1 \varepsilon(\theta+u\delta)du\right]^{1/2} \to 0$$

cuando $\delta \to 0$.

Así pues, hemos demostrado (4) suponiendo que $I_n(\Theta)$ es continua. Pero para $n = 1$, $I_n(\theta) = I(\theta)$, esta suposición también se cumple en virtud de las condiciones (R). Por lo tanto, (4) es justa cuando $n = 1$ y, por consiguiente, también es justa (5). Pero de (5) resulta la relación (6) que significa la continuidad de $I(\theta)$. El teorema queda demostrado.

Teorema 3. *Si el conjunto Θ es compacto y la función $\sqrt{f_\theta(x)}$ para $[\mu]$ c.t. valores de x es continuamente derivable respecto a θ, entonces, la continuidad de $I(\theta)$ tendrá lugar si y sólo si se cumple* (3).

El teorema significa que la continuidad de $I(\theta)$ en la condición (R) puede ser sustituida por la condición (3).

Demostración. Supongamos que $f(\theta)$ se continua y que no se cumple (3). Entonces existe $\gamma > 0$, y las sucesiones $t \to \theta \in \Theta$ y $N_t \to \infty$ son tales que

$$m(t) \equiv \mathbf{M}_t[|l'(x_1, t)|^2;\ l'(x_1, t) > N_t] > \gamma \tag{9}$$

para todos los valores de t de la sucesión elegida.

Utilicemos el teorema 1 para $\mathscr{Y} = \mathscr{X}$, $\nu = \mu$, $\psi(t, x) = (\sqrt{f_t(x)}')^2 = \frac{1}{4}(l'(x, t))^2 f_t(x)$, $A(t) = \{x;\ |\sqrt{f_t(x)}'| \leqslant 2|\sqrt{f_\theta(x)}'|\}$. En virtud de la continuidad de $\sqrt{f_\theta(x)}'$, las condiciones 1) y 2) del teorema 1 se cumplen y, por consiguiente, de la continuidad de $I(t)$ se deducirá que

$$m_1(t) = \int\limits_{x\notin A(t)} |\sqrt{f_t(x)}'|^2\mu(dx) \to 0$$

cuando $t \to 0$. Pero $m(t) \leqslant m_1(t) + m_2(t)$, donde

$$m_2(t) = \int\limits_{B(t)\cap A(t)} (\sqrt{f_t}')^2\mu,\quad B(t) = \{x;\ 2|\sqrt{f_1(x)}'| > N_t\sqrt{f_1(x)}\}.$$

De la forma del conjunto $A(t)$ resulta

$$m_2(t) \leqslant 4\int\limits_{B(t)} |\sqrt{f_\theta}'|^2\mu.$$

Volviendo a utilizar la convergencia $(\sqrt{f_t(x)})' \to (\sqrt{f_\theta(x)})'$, $\sqrt{f_t(x)} \to \sqrt{f_\theta(x)}$ para $t \to \theta$, obtenemos que $B(t)$ converge hacia el conjunto de μ-medida 0. Esto significa que $\mu(B(t) \to 0$, $m_2(t) \to 0$, $m(t) \to 0$ cuando $t \to \infty$. Hemos obtenido la contradicción con (9). La relación (3) queda demostrada.

Ahora supongamos que se cumple (3). En virtud del teorema 1, para demostrar la continuidad $I(t)$ es suficiente convencerse que con el mismo conjunto $A(t)$ que hemos utilizado más arriba, se cumple $m_1(t) \to 0$ cuando $t \to \infty$. Pero

$$m_1(t) \leqslant \int\limits_{|l'|>N} |l'|^2 f_t\mu + N^2 \int\limits_{x \notin A(t)} f_t\mu.$$

donde, por medio de la elección de N, la primera integral puede hacerse, en virtud de (3), tan pequeña como se quiera. Para estimar la segunda integral es necesario notar que $\mu(A(t)) \to 0$ y que cuando $C(t) = \{x; f_t(x) \leqslant 2f_\theta(x)\}$ se cumple $\int\limits_{x \notin C(t)} f_t\mu \to 0$ cuando $t \to 0$ (véase la demostración del corolario 1). Por eso

$$\int\limits_{x \notin A(t)} f_t\mu \leqslant 2 \int\limits_{x \notin A(t)} f_\theta\mu + \int\limits_{x \notin C(t)} f_t\mu \to 0$$

cuando $t \to 0$. ◁

3. Corolarios de las condiciones(RR).

Teorema 4. *Si se cumplen las condiciones (RR), entonces* $\int f_\theta''(x)\mu(dx) = 0$.

Junto con el teorema 2 esto asegura el cumplimiento de las condiciones (2.24.4) que necesitamos en el § 2.24.

Demostración. En virtud del teorema 2, para todos $\theta \in \Theta$,

$$\int f_\theta'(x)\mu(dx) = 0$$

y nos es suficiente demostrar que, cuando $t \to 0$,

$$J(t) = \frac{1}{t-\theta}\left[\int f_t'\mu - \int f_\theta'\mu\right] \to \int f_\theta''\mu.$$

Nótese que $\frac{1}{t-\theta}(f_t' - f_\theta') = \varphi_t f_t + \frac{f_\theta'}{f_\theta}\cdot\frac{f_t - f_\theta}{t-\theta}$, donde $\varphi_t = \frac{1}{t-\theta}\left(\frac{f_t'}{f_t} - \frac{f_\theta'}{f_\theta}\right)$. Aprovechando esta igualdad podemos representar $J(t)$ en forma de la suma de cuatro sumandos: $J(t) = J_1 + J_2 + J_3 + J_4$, donde

$$J_1 = \int \varphi_t f_\theta\mu, \quad J_2 = \int\limits_{l \leqslant N} \varphi_t(f_t - f_\theta)\mu,$$

$$J_3 = \int\limits_{l \geqslant N} \varphi_t(f_t - f_\theta)\mu, \quad J_4 = \int \frac{f_\theta'}{f_\theta}\cdot\frac{f_t - f_\theta}{t-\theta}\,\mu,$$

$l = l(x)$ es la mayorante para $l''(x, t)$ en las condiciones (RR). En virtud del teorema 2, cuando $n = 1$, $S(x) = l'(x, \theta)$ obtenemos

$$J_4 = \frac{1}{t-\theta}(\mathbf{M}_t l'(x_1, \theta) - \mathbf{M}_\theta l'(x_1, \theta)) \to \mathbf{M}_\theta(l'(x_1, \theta))^2 = I(\theta). \tag{10}$$

Seguidamente,

$$|\varphi_t| < l \tag{11}$$

y, por lo tanto, según el teorema de Lebesgue,

$$\lim_{t \to \theta} J_1 = \int \lim_{t \to \theta} \varphi_t f_\theta\mu = \int l'' f_\theta\mu = \int f_\theta''\mu - I(\theta). \tag{12}$$

Volviendo a utilizar (11), obtenemos, en virtud de las condiciones (RR),

$$|J_3| \leqslant \int_{l \geqslant N} lf_t\mu + \int_{l \geqslant N} lf_\theta\mu \to 0$$

cuando $N \to \infty$. Por último, en virtud de la desigualdad de Cauchy — Buniakovski,

$$|J_2| \leqslant N\int |f_t - f_\theta|\mu \leqslant N\int \int_0^t |f'_u| du\mu \leqslant N \int_0^1 \sqrt{I(u)}du \to 0 \tag{13}$$

cuando $t \to \theta$. Comparando (10)—(13) obtenemos que $0 \equiv J(t) \to \int f''_\theta\mu$. ◁

Suplemento VII

Desigualdades para la distribución de la relación de verosimilitud en el caso multidimensional

En este apartado demostraremos el siguiente teorema (teorema 28.2; las designaciones véanse en los §§ 2.21, 2.23, 2.28).

Teorema 1. *Supongamos que se cumplen las condiciones siguientes:*

$$\inf_u \frac{r(u)}{u^2} \geqslant g > 0, \tag{1}$$

$$\mathbf{M}_\theta l'(x_1, \theta) = 0, \tag{2}$$

$$\gamma = \sup_\theta \mathbf{M}_\theta |l'(x_1, \theta)|^s > \infty \tag{3}$$

para cierto $s > k$. Entonces, para cualesquiera z, $n \geqslant 1$, $r > 0$,

$$\mathbf{P}_\theta \left(\sup_{|v| \geqslant r} Z\left(\frac{v}{\sqrt{n}}\right) \geqslant e^z \right) \leqslant c\gamma(e^{-z} + e^{-z/2})e^{-\beta g r^2},$$

donde $\beta > 0$ depende únicamente de k y s, $c < \infty$ depende de k, s y g.

Como ya hemos señalado en el § 2.28, para demostrar este teorema utilizaremos la posibilidad de estimar $\sup\limits_{u \in K_{0,1}} p(u)$ para cierta función p y para el cubo unitario

$$K_{0,1} = \{u = (u_1, \ldots, u_k)\colon 0 \leqslant u_j \leqslant 1, \ \ j = 1, \ldots, k\}$$

a través de los valores de $p(0)$ y $\int\limits_{K_{0,1}} |p'(u)|^s du (p'(u) = \operatorname{grad} p(u)$. Para realizar esta posibilidad necesitaremos la siguiente afirmación, cuya demostración reproducimos aquí, puesto que no figura en los conocidos manuales de análisis matemático. Por C_k, C_s y $C_{k,s}$ designaremos distintas constantes que sólo dependen de sus índices.

Lema 1. *Para cualquier $s > k$ existe $c_{k,s}$ tal, que*

$$\sup_{x \in K_{0,1}} |p(x)| \leqslant |p(x)| + c_{k,s}\left[\int_{K_{0,1}} |p'(x)|^s dx\right]^{1/s}$$

para cualquier $x \in K_{0,1}$.

Demostración. Para x, $y \in k_{0,1}$ es válida

$$p(x) = p(y) + \int_0^1 (p'(y + t(x - y)), x - y)dt.$$

Integrando esta igualdad respecto a $y \in K_{0,1}$, obtenemos

$$p(x) = \int\limits_{K_{0,1}} p(y)dy + \int\limits_{K_{0,1}} \int\limits_0^1 (p'(y + t(x - y)), \ \ x - y)dtdy \equiv I_1 + I_2, \tag{4}$$

donde I_1, I_2 designan el primero y el segundo, respectivamente. Sustituyamos en la integral I_2, las variables $y = \dfrac{z - tx}{1 - t}$. Entonces

$$y + t(x - y) = z, \quad x - y = \frac{x - z}{1 - t}, \quad dy = \frac{dz}{(1 - t)^k},$$

$$I_2 = \int\limits_{K_{0,1}} (p'(z), x - z)K(x, z)dz, \tag{5}$$

donde $K(x, z) = \displaystyle\int\limits_0^1 \varphi\left(\frac{z - tx}{1 - t}\right) \frac{dt}{(1 - t)^{k+1}}$, φ es el indicador del cubo $K_{0,1}$. Si aquí sustituimos $t = 1 - \dfrac{|z - x|}{u}$, entonces $\dfrac{1}{1 - t} = \dfrac{u}{|z - x|}$ y podemos escribir

$$K(x, z) = |z - x|^{-k} \int\limits_{|z-x|}^{\infty} \varphi\left(x + \frac{z - x}{|z - x|} u\right) u^{k-1}du.$$

En vista de que para cualesquiera z, x, el portador de la función $\varphi\left(x + \dfrac{z - x}{|z - x|} u\right)$ está presente en el segmento $[0, 2\sqrt{k}]$, entonces

$$K(x, z) \leqslant |z - x|^{-k} \int\limits_0^{2\sqrt{k}} u^{k-1}du = \frac{(2\sqrt{k})^k}{k|z - x|^k}.$$

Utilizando (5) y la desigualdad de Hölder, obtenemos

$$|I_2| \leqslant C_k \int\limits_{K_{0,1}} \frac{|p'(z)|}{|z - x|^{k-1}} dz \leqslant c_k J \left(\int\limits_{K_{0,1}} |p'(z)|^s dz \right)^{1/s},$$

donde

$$c_k = k^{-1}(2\sqrt{k})^k, \quad J = \left(\int\limits_{K_{0,1}} \frac{dz}{|z - x|^{(k-1)r}} \right)^{1/r}, \ \frac{1}{s} + \frac{1}{r} = 1.$$

Pero cuando $s > k$ se cumple $(k - 1)r = (k - 1)\dfrac{s}{s - 1} < k$,

$$J = J(k, s, x) < \left(\int\limits_K \frac{dz}{|z|^{(k-1)r}} \right)^{1/r} \equiv J(k, s) < \infty,$$

donde k es un cubo, o sea, $K = \{z\colon |z_j| \leqslant 1, \ j = 1, \ldots, k\}$.

Ahora bien, en virtud de (4),

$$\sup_{x\in K_{0,1}} |p(x)| \leqslant |I_1| + \sup_{x\in K_{0,1}} |I_2| \leqslant |p(x)| + 2\sup_{x\in K_{0,1}} |I_2| \leqslant$$
$$\leqslant p(x) + 2c_k I(k, s)\left(\int_{K_{0,1}} |p'(z)|^s dz\right)^{1/s}.$$

El lema queda demostrado.

Así pues, la estimación de $\sup_{x\in K_{0,1}} |p(x)|$ es posible en los términos de $|p(x)|$ cuando está fijo $x \in K_{0,1}$ y $\int_{K_{0,1}} |p'(u)|^s du$ para $s > k$. Si seguimos el método que hemos utilizado en el caso unidimensional, ahora necesitaremos estimar $\mathbf{M}_\theta |p'(u)|^s$, donde, en calidad de $p(u)$ elegiremos la función

$$p(u) = Z^{1/s}(u). \tag{6}$$

Para esto, a su vez, necesitaremos los lemas siguientes.

Lema 2. *Sean* ξ_j, $j = 1, 2, \ldots$, *los vectores independientes e igualmente distribuidos de* R^k, $\mathbf{M}\xi_1 = 0$, $\mathbf{M}|\xi_1|^s \leqslant \gamma < \infty$, $s \geqslant 2$. *Entonces*

$$\mathbf{M}\left|\sum_{j=1}^n \xi_j\right|^s \leqslant c_{k,s}\gamma n^{s/2}.$$

Demostración. Para simplificar los razonamientos nos limitaremos a examinar el caso cuando $s = 2m$ es un número entero par*. En este caso es suficiente examinar las variables aleatorias escalares ξ_j, puesto que $\xi_j = (\xi_{j,1}, \ldots, \xi_{j,k})$,

$$\left|\sum_{j=1}^n \xi_j\right|^{2m} = \left[\left(\sum_{j=1}^n \xi_{j,1}\right)^2 + \ldots + \left(\sum_{j=1}^n \xi_{j,k}\right)^2\right]^m$$

y, en virtud de la desigualdad de Minkovski,

$$\left(\mathbf{M}\left|\sum_{j=1}^n \xi_j\right|^{2m}\right)^{1/m} \leqslant \left[\mathbf{M}\left(\sum_{j=1}^n \xi_{j,1}\right)^{2m}\right]^{1/m} + \ldots + \left[\mathbf{M}\left(\sum_{j=1}^n \xi_{j,k}\right)^{2m}\right]^{1/m}.$$

Para las ξ_j escalares tenemos

$$\mathbf{M}\left|\sum_{j=1}^n \xi_j\right|^s = \sum_{j_1,\ldots,j_n} \mathbf{M}\xi_1^{j_1}\ldots\xi_n^{j_n}, \tag{7}$$

donde la suma se realiza con arreglo a todos $j_1, \ldots, j_n$ enteros, tales, que $\sum_l j_l = s$, $j_l \neq 1$ ($j_l = 1$ se excluyen, ya que $\mathbf{M}\xi_j = 0$). Según la desigualdad de Hölder,

$$|\mathbf{M}\xi_l^{j_l}| \leqslant (\mathbf{M}|\xi_l|^s)^{j_l/s} = \gamma^{j_l/s}$$

y, por consiguiente,

$$\prod_{l=1}^n |\mathbf{M}\xi_l^{j_l}| \leqslant \prod_{l=1}^n \gamma^{j_l/s} = \gamma.$$

Nos queda estimar $\sum_{j_1,\ldots,j_n} 1$. Designemos por $(k_1, \ldots, k_p)$ los elementos no nulos ($k_l \geqslant 2$) del conjunto $(j_1, \ldots, j_n)$ $\left(\sum_{l=1}^p k_l = s\right)$. Entonces, la suma sujeta a estimación será igual a

---

*) La demostración en el caso general véase, por ejemplo, en [31], p. 255.

$\sum\limits_{(k_1,\ldots,k_p)} A_p$, donde A_p es el número de ubicaciones de los elementos $k_1, \ldots, k_p$ en n lugares. Es evidente que $A_p \leqslant n(n-1)\ldots(n-p+1)$. El valor mayor posible de p es igual a $m = s/2$ (éste corresponde al conjunto $(2, 2, \ldots, 2)$), así que $A_p \leqslant A_m \leqslant n^m$. Pero el número de conjuntos diferentes $(k_1, \ldots, k_p)$ depende exclusivamente de s. Por consiguiente, la suma estimada no supera $c_s n^m$. ◁

Supongamos que la función $p(u)$ ha sido definida en (6).

Lema 3. *Si se cumplen las condiciones* (2) y (3),

$$\mathbf{M}_\theta |p'(u)|^s \leqslant c_s \gamma n^{s/2}.$$

Demostración

$$\mathbf{M}_\theta |p'(u)|^s = \mathbf{M}_\theta \left| \frac{1}{s} L'(X, \theta + u) Z^{1/s}(u) \right|^s =$$
$$= s^{-s}\mathbf{M}_\theta |L'(X, \theta + u)|^s Z(u) = s^{-s}\mathbf{M}_{\theta+u} |L'(X, \theta + u)|^s.$$

Nos queda utilizar el lema 2, aplicándolo a las variables aleatorias $\xi_j = l'(x_j, \theta + u)$.

Designemos por $K_{u,\Delta}$ el cubo en R^k, con lado de longitud Δ y con vértice en el punto $u = (u_1, \ldots, u_k)$:

$$K_{u,\Delta} = \{v \in R^k \colon u_i \leqslant v_i \leqslant u_i + \Delta,\ i = 1, \ldots, k\}.$$

Lema 4. *Si se cumplen las condiciones del teorema 1,*

$$\mathbf{P}_\theta \left(\sup_{v \in K_{u,\Delta}} Z\left(\frac{v}{\sqrt{n}}\right) > e^z \right) \leqslant c_{k,s}\gamma\Delta^k (e^{-z} + e^{-z/2}) e^{-|u|^2 \beta g},$$

donde $\beta = \min\left(\frac{1}{4}, \frac{s-k}{4k}\right)$, $\Delta = e^{-|u|^2 g/4k}$.

Esta misma estimación será cierta para cualquier cubo con lado de longitud Δ *y que contiene el punto* u.

Demostración. Representemos el punto $v \in K_{u,\Delta}$ en forma de $v = u + t\Delta$, donde $t \in K_{0,1}$. Entonces

$$\mathbf{P} = \mathbf{P}_\theta \left(\sup_{v \in K_{u,\Delta}} Z\left(\frac{v}{\sqrt{n}}\right) > e^z \right) = \mathbf{P}_\theta \left(\sup_{t \in K_{0,1}} Z^{1/s}\left(\frac{u + t\Delta}{\sqrt{n}}\right) \geqslant e^{z/s} \right) =$$
$$= \mathbf{P}_\theta \left(\sup_{t \in K_{0,1}} p\left(\frac{u + t\Delta}{\sqrt{n}}\right) \geqslant e^{z/s} \right).$$

En virtud del lema 1,

$$\mathbf{P} \leqslant \mathbf{P}_\theta \left(p\left(\frac{u}{\sqrt{n}}\right) \geqslant \frac{1}{2} e^{z/s} \right) + \mathbf{P}_\theta \left(\left[\int_{K_{0,1}} \left| p_t'\left(\frac{u + t\Delta}{\sqrt{n}}\right) \right|^s dt \right]^{1/s} \geqslant \right.$$
$$\left. \geqslant \frac{1}{2c_{k,s}} e^{z/s} \right) = \mathbf{P}_{(1)} + \mathbf{P}_{(2)},$$

donde $\mathbf{P}_{(1)}$ y $\mathbf{P}_{(2)}$ designan el primero y el segundo sumandos, respectivamente. Estimemos $\mathbf{P}_{(1)}$ con ayuda de la desigualdad de Chébishev y del teorema 28.1:

$$\mathbf{P}_{(1)} \leqslant 2^{s/2} e^{-z/2} \mathbf{M}_\theta Z^{1/2}\left(\frac{u}{\sqrt{n}}\right) \leqslant 2^{s/2} e^{-z/2} e^{-\frac{|u^2| g}{2}}. \tag{8}$$

Para estimar $\mathbf{P}_2$ también utilizaremos la desigualdad de Chébishev:

$$\mathbf{P}_{(2)} = \mathbf{P}_\theta\left(\int\limits_{K_{0,1}}\left(\frac{\Delta}{\sqrt{n}}\right)^s\left|p'\left(\frac{u+t\Delta}{\sqrt{n}}\right)\right|^s dt \geqslant \left[\frac{1}{2c_{k,s}}\right]^s e^z\right) \leqslant$$

$$\leqslant e^{-z}(2c_{k,s})^s \mathbf{M}_\theta \int\limits_{K_{0,1}}\left(\frac{\Delta}{\sqrt{n}}\right)^s\left|p'\left(\frac{u+t\Delta}{\sqrt{n}}\right)\right|^s dt =$$

$$= (2c_{k,s})^s e^{-z}\left(\frac{\Delta}{\sqrt{n}}\right)^s \int\limits_{K_{0,1}} \mathbf{M}_\theta\left|p'\left(\frac{u+t\Delta}{\sqrt{n}}\right)\right|^s dt.$$

En virtud del lema 3,

$$\mathbf{P}_{(2)} \leqslant c_{k,s}e^{-z}\left(\frac{\Delta}{\sqrt{n}}\right)^s \gamma n^{s/2} = c_{k,s}\gamma e^{-z}\Delta^s. \tag{9}$$

Poniendo

$$\Delta = e^{-\frac{|u|^2 g}{4k}}$$

y suponiendo, sin limitar la generalidad, $\gamma \geqslant 1$, obtenemos

$$\mathbf{P} \leqslant 2^{s/2}e^{-z/2}e^{-\frac{|u|^2 g}{2}} + c_{k,s}\gamma e^{-z}\Delta^s \leqslant (2^{s/2} + c_{k,s})\gamma\Delta^k \times$$

$$\times\left[e^{-z/2}e^{-\frac{|u|^2 g}{4}} + e^{-z}e^{-\frac{|u|^2 g^{(s-k)}}{4k}}\right] \leqslant (2^{s/2} + c_{k,s})\gamma\Delta^k(e^{-z/2} + e^{-z})e^{-\beta g|u|^2},$$

donde

$$\beta = \min\left(\frac{1}{4}, \frac{s-k}{4k}\right).$$

La última afirmación del lema se deduce, evidentemente, del lema 1 y de la demostración expuesta. El lema queda demostrado. ◁

Demostración del teorema 1. Cubramos todo el espacio R^k de un sistema de cubos $K_{u,\Delta}$ en los que las coordenadas de los puntos u son múltiplos de Δ. El número de tales cubos, que se intersecan con la capa $S_r = \{v \in R^k: r \leqslant |v| \leqslant r+1\}$, está limitado por la cantidad $c_k r^{k-1}$. Por lo tanto,

$$\mathbf{P}_\theta\left(\sup_{v\in S_r} Z\left(\frac{v}{\sqrt{n}}\right) \geqslant e^z\right) \leqslant c_k r^{k-1} c_{k,s}\gamma(e^{-z/2} + e^{-z})e^{-r^2\beta g},$$

$$\mathbf{P}_\theta\left(\sup_{|v|\geqslant r} Z\left(\frac{v}{\sqrt{n}}\right) \geqslant e^z\right) \leqslant c_k c_{k,s}\gamma(e^{-z/2} + e^{-z}) \sum_{j=0}^{\infty} (r+j)^{k-1}e^{-(r+j)\beta g}.$$

La sucesión $(r+j)^{k-1}e^{-(r+j)^2\beta g}$ para todos $j \geqslant j(k, \beta g)$, donde $j(k, \beta g)$ depende únicamente de sus argumentos, decrece más rápidamente que la progresión geométrica con exponente $\frac{1}{2}$. Por eso, la serie en el segundo miembro de (10) no supera, para todos r, el primer sumando con una exactitud de hasta la constante que sólo depende de k y βg. Como $\sup\limits_{r \geqslant 0} r^{k-1} \times$ $\times\ e^{-\frac{1}{2}r^2\beta g} < \infty$ también depende únicamente de k y βg, entonces

$$\mathbf{P}_\theta \left(\sup_{|v| \geqslant r} Z \left(\frac{v}{\sqrt{n}} \right) > e^z \right) \leqslant c\gamma(e^{-z/2} + e^{-z}) e^{-\frac{1}{2}r^2\beta g},$$

donde c depende de k, s y βg. Sustituyendo aquí $\frac{\beta}{2}$ por β, obtenemos la afirmación del teorema. ◁

Suplemento VIII

Demostración de dos teoremas fundamentales de la teoría de los juegos estadísticos

Aquí vamos a suponer que se cumplen las condiciones siguientes.

Condición (A). *El conjunto de decisiones D y el conjunto de parámetros (estrategias puras de la naturaleza) Θ son espacios métricos compactos con métricas ϱ_D y ϱ_Θ, respectivamente.*

Condición (B). *La función de pérdidas $w(\delta, \theta)$: $D \times \Theta \to R$ es continua respecto a δ y θ en las métricas ϱ_D y ϱ_Θ, respectivamente.*

No necesitaremos la propiedad de $w(\delta, \theta) \geqslant 0$ y no supondremos que ésta tenga lugar.

Además, disponemos de la muestra $X \in \mathbf{P}_\theta$ de la distribución $\mathbf{P}_\theta$. Su volumen n, sin limitar la generalidad, se puede considerar igual a 1.

Condición (C). *Las distribuciones $\mathbf{P}_\theta$, con arreglo a la variación son continuas respecto a θ, o sea,*

$$\sup_{B \in \mathfrak{B}_{\mathscr{X}}} |\mathbf{P}_{\theta_m}(B) - \mathbf{P}_\theta(B)| \to 0$$

si $\varrho_\Theta(\theta_m, \theta) \to 0$ *cuando* $m \to \infty$.

Si se cumple la condición (A$_\mu$), o sea, si $\mathbf{P}_\theta$ tiene una densidad $f_\theta(x)$ respecto a cierta medida σ-finita μ en $(\mathscr{X}, \mathfrak{B}_{\mathscr{X}})$:

$$f_\theta(x) = \frac{d\mathbf{P}_\theta}{d_\mu}(x),$$

entonces la condición (C) será equivalente a la continuidad de $f_\theta(x)$ en $L_1(\mathscr{X}, \mathfrak{B}_{\mathscr{X}}, \mu)$:

$$\int |f_{\theta_m}(x) - f_\theta(x)| \mu(dx) \to 0$$

si $\varrho_\Theta(\theta_m, \theta) \to 0$ cuando $m \to \infty$.

Las condiciones (A), (B) y (C) admiten, claro está, la posibilidad de ser finitas a los conjuntos D y Θ.

Si D es finito y consta de los puntos $\delta_1, \ldots, \delta_r$, entonces se cumplirá la condición A respecto a D (la elección de ϱ_D no tiene importancia), y la condición (B) significará la continuidad de las funciones $w(\delta_1, \theta), \ldots, w(\delta_r, \theta)$ respecto a ϱ_Θ.

Si ambos conjuntos D y Θ son finitos, las condiciones (A), (B) y (C) serán cumplidas automáticamente.

Designemos por σ_D y σ_Θ las σ-álgebras de los conjuntos de Borel de D y de Θ, respectivamente. Siguiendo el § 5.3, designemos por $(\tilde{\mathscr{D}}, \tilde{\Theta}, \tilde{W})$ el juego estadístico promediado, donde como elementos de $\tilde{\Theta}$ sirven las distribuciones $\mathbf{Q}$ en (Θ, σ_Θ), y como elementos de $\tilde{\mathscr{D}}$, las distribuciones $\pi(x) = \pi(x, \cdot)$ en (D, σ_D) (para cada $x \in \mathscr{X}$), donde $\pi(x, A)$ para cada $A \in \sigma_D$ es una función medible respecto a x.

La función de riesgo $\tilde{W}(\pi, \mathbf{Q})$ es definida por la igualdad

$$\tilde{W}(\pi, \mathbf{Q}) = \int_\Theta \int_{\mathscr{X}} \int_D w(u, t)\pi(x, du) f_t(x)\mu(dx)\mathbf{Q}(dt).$$

Si en vez del argumento $\mathbf{Q}$ se pone θ, entonces $\tilde{W}(\pi, \theta)$ significará $\tilde{W}(\pi, I_\theta)$, donde I_θ es la distribución concentrada en el punto θ. Este mismo acuerdo será válido respecto a la sustitución de $\pi \in \tilde{\mathscr{D}}$, por $\delta \in \mathscr{D}$. También será más cómodo escribir W en vez de $\tilde{W}$, ya que esto nunca conducirá a equivocaciones.

Lema 1. *Si se cumplen las condiciones* (A), (B), (C), *la función* $W(\pi, \theta)$ *será continua en* θ *para cualquier estrategia* $\pi(x)$.

Demostración. Tenemos para $\theta_n \to \theta$:

$$\begin{aligned} |W(\pi, \theta_n) - W(\pi, \theta)| &\leqslant |\mathbf{M}_\theta \mathbf{M}[w(\pi(X), \theta) - w(\pi(X), \theta_n)/X]| + \\ &+ |\mathbf{M}_\theta \mathbf{M}[w(\pi(X), \theta_n)/X] - \mathbf{M}_{\theta_n}\mathbf{M}[w(\pi(X), \theta_n)/X]| \leqslant \\ &\leqslant \int |w(\pi(x), \theta) - w(\pi(x), \theta_n)| \mathbf{P}_\theta(dx) + \sup_{\delta, \theta} |w(\delta, \theta)| \int |\mathbf{P}_{\theta_n}(dx) - \mathbf{P}_\theta(dx)|. \end{aligned} \tag{1}$$

La primera integral aquí converge a 0 en virtud de la continuidad de la función w respecto a θ. La convergencia a cero de la segunda integral se deduce de la condición (C). En efecto, sea $f_{\theta_n}(x)$ la densidad $\mathbf{P}_{\theta_n}$ respecto a la medida

$$\mu = \mathbf{P}_\theta + \sum_{j=1}^{\infty} 2^{-j}\mathbf{P}_{\theta_j},$$

y sea $B_n = \{x: f_{\theta_n}(x) \geqslant f_\theta(x)\}$. Entonces, la segunda integral en (1) será igual a

$$\int |f_{\theta_n}(x) - f_\theta(x)|\mu(dx) = 2\int_{B_n} (f_{\theta_n}(x) - f_\theta(x))\mu(dx) = 2(\mathbf{P}_{\theta_n}(B_n) - \mathbf{P}_\theta(B_n)) \to 0.$$

El lema queda demostrado.

Teorema 1. (primer teorema fundamental). *Si se cumplen las condiciones* (A), (B) y (C), *el juego* $(\tilde{\mathscr{D}}, \tilde{\Theta}, W)$ *tendrá precio y estrategias minimáx de ambos jugadores. Con otras palabras, existirá la distribución menos favorable* $\overline{\mathbf{Q}}$ *y la regla minimáx de decisión* $\overline{\pi}(x)$:

$$W_* = \sup_{\mathbf{Q}} \inf_{\pi} W(\pi, \mathbf{Q}) = W(\overline{\pi}, \overline{\mathbf{Q}}) = \inf_{\pi} \sup_{\mathbf{Q}} W(\pi, \mathbf{Q}) = W^*. \tag{2}$$

En virtud del lema 2.1, la afirmación (2) es equivalente al hecho de que

$$W(\overline{\pi}, \uparrow) \equiv \sup_{\mathbf{Q}} W(\overline{\pi}, \overline{\mathbf{Q}}) = W(\overline{\pi}, \overline{\mathbf{Q}}) = \inf_{\pi} W(\pi, \overline{\mathbf{Q}}) \equiv W(\downarrow, \overline{\mathbf{Q}}). \tag{3}$$

Teorema 2 (segundo teorema fundamental). *Si se cumplen las condiciones* (A), (B) y (C), *las decisiones bayesianas* $\pi_Q(x)$ *formarán una clase completa. Con otras palabras, para cualquier* $\pi_0 \in \tilde{\mathscr{D}}$ *habrá* $\mathbf{Q} \in \tilde{\Theta}$, $\pi_Q \in \tilde{\mathscr{D}}$ *tales, que*

1) $W(\pi_Q, \mathbf{Q}) = W(\downarrow, \mathbf{Q})$,

2) $W(\pi_Q, \theta) \leqslant W(\pi_0, \theta)$ *para todos* θ.

Demostración del teorema 2. El segundo teorema fundamental es el corolario del primero. Examinemos la estrategia arbitraria $\pi_0 \in \tilde{\mathscr{D}}$ y el juego $(\tilde{\mathscr{D}}, \tilde{\Theta}, W_0)$, donde W_0 se ha construido a base de la función $w_0(\delta, \theta) = w(\delta, \theta) - W(\pi_0, \theta)$, así que

$$W_0(\pi, \theta) = W(\pi, \theta) - W(\pi_0, \theta). \tag{4}$$

En virtud del lema 1, la función $v(\theta) = W(\pi_0, \theta)$ es continua en θ y, por lo tanto, la función de pérdidas $w_0(\delta, \theta) = w(\delta, \theta) - v(\theta)$, junto con $w(\delta, \theta)$, satisface la condición (B). Esto significa que el teorema 1 es aplicable al juego $(\mathscr{D}, \tilde{\Theta}, W_0)$. En vista de que $W_0(\pi_0, \uparrow) = 0$ (véase (4)), el precio superior de este juego satisface la condición $W_0^* \leqslant 0$. Entonces, de (2) y (3) se deduce que existen $\bar{\pi}$, $\overline{\mathbf{Q}}$ tales, que

$$\sup_{\mathbf{P}} W_0(\bar{\pi}, \mathbf{P}) = \sup_{\theta} W_0(\bar{\pi}, \theta) \leqslant 0, \; \bar{\pi} = \pi_{\overline{Q}}.$$

Estas dos relaciones son equivalentes a las afirmaciones 2) y 1) del teorema 2 si se pone $\overline{\mathbf{Q}} = \mathbf{Q}$, $\bar{\pi} = \pi_Q$. El teorema queda demostrado.

La demostración del teorema 1 se deducirá de los dos lemas siguientes.

Lema 2. *Al cumplirse las condiciones* (A), (B) y (C) *existirá una distribución* $\overline{\mathbf{Q}}$ *tal, que* $W(\downarrow, \overline{\mathbf{Q}}) \geqslant \inf W(\pi, \uparrow) = W^*$.

Lema 3. *Al cumplirse las condiciones* (A), (B) y (C) *existirá una estrategia* $\bar{\pi}$ *tal, que* $W(\bar{\pi}, \uparrow) \leqslant W^*$.

De las desigualdades de los lemas 2 y 3 se desprende la relación

$$W^* \geqslant W(\bar{\pi}, \uparrow) \geqslant W(\bar{\pi}, \overline{\mathbf{Q}}) \geqslant W(\downarrow, \overline{\mathbf{Q}}) \geqslant W^*$$

equivalente a (3) y, por consiguiente, a (2). Esto demuestra el teorema 1. ◁

Los lemas 2 y 3 dividen la demostración del teorema 1 en dos partes. La primera de ellas (lema 2) está muy poco relacionada con el hecho de que el juego es estadístico. Esta parte de la demostración se realiza aproximadamente igual que para los juegos ordinarios (compárese con [31]).

Demostración del lema 2. Sea V un conjunto de funciones $\Theta \to R$ representables en forma de $v(\theta) = W(\pi, \theta)$, $\pi \in \mathscr{D}$. En virtud del lema 1, todas las funciones de V son continuas, así que $V \subset C(\Theta)$, donde $C(\Theta)$ es el espacio de todas las funciones continuas en Θ. Asimismo, sea $v_1(\theta) = W(\pi_1, \theta)$, $v_2(\theta) = W(\pi_2, \theta)$. En vista de que para $p \in [0, 1]$,

$$v(\theta) = pv_1(\theta) + (1 - p)v_2(\theta) = W(p\pi_1 + (1 - p)\pi_2, \theta),$$
$$\pi = p\pi_1 + (1 - p)\pi_2 \in \mathscr{D},$$

entonces, $v \in V$ y, por lo tanto, el conjunto V es convexo.

Ahora notemos que $W^* = \inf_{\pi} W(\pi, \uparrow) = \inf_{v \in V} \sup_{\theta} v(\theta)$. En vez de la función inicial $w(\delta, \theta)$ no será más cómodo examinar la función $\dfrac{w(\delta, \theta) - v_0 + 1}{W^* - v_0 + 1}$, $v_0 = \inf_{v \in V} \inf_{\theta} v(\theta)$. Designando la nueva función otra vez por $w(\delta, \theta)$ (en este caso el problema queda invariable), obtenemos que para ella

$$W^* = 1, \quad v_0 > 0. \tag{5}$$

Sea ahora U un conjunto de funciones continuas $v(\theta)$: $\Theta \to R$ tales, que $\sup_{\theta} v(\theta) < 1$. Es evidente que U es un conjunto abierto convexo de $C(\Theta)$. Además, de (5) se deduce que la intersección $V \cap U$ está vacía. Por eso, en virtud del teorema de Hahn — Banach (véase, por ejemplo, [31], p. 171, 200—206) existe una funcional lineal $L(v)$: $C(\theta) \to R$ tal, que

$$L(v) < 1 \quad \text{para} \quad v \in U, \quad L(v) \geqslant 1 \quad \text{para } v \in V. \tag{6}$$

Esta funcional posee, cuando es necesario, la propiedad $L(v) \geqslant 0$ si $v(\downarrow) = \inf_{\theta} v(\theta) \geqslant 0$. En efecto, admitiendo la existencia del elemento $v_0 \in C(\Theta)$, $v_0(\downarrow) \geqslant 0$, para el cual $L(v_0) < 0$, obtenemos que $v_s = -sv_0 \in U$, cualquiera que sea $s > 0$, $L(v_s) = -sL(v_0) > 1$ y siempre que s sea bastante grande. Esto conduce a cierta contradicción con (6).

Pero la funcional no negativa L, en virtud del teorema de Riesz ([42)], p. 240), admite la representación en forma de la integral

$$L(v) = \int_{\Theta} v(\theta)\lambda(d\theta),$$

donde λ es una medida finita. Como $1 \geqslant \sup_{v \in U} L(v) = \lambda(\Theta)$, entonces, poniendo $\overline{\mathbf{Q}}(A) = \lambda(A)/\lambda(\Theta)$, obtenemos para $v \in V$:

$$L(v) = \int W(\pi, \theta)\lambda(d\theta) = \lambda(\Theta) W(\pi, \overline{\mathbf{Q}}),$$

$$W(\downarrow, \overline{\mathbf{Q}}) = \frac{1}{\lambda(\Theta)} \inf_{v \in V} L(v) \geqslant 1 = W^*.$$

El lema queda demostrado.

Demostración del lema 3. En vista de que la función $W(\pi, \Theta)$ para cada $\pi \in \mathscr{D}$ es continua respecto a θ (véase el lema 1), nos es suficiente construir la estrategia $\overline{\pi}$ para la cual, con todos $k = 1, 2, \ldots$,

$$W(\overline{\pi}, \theta_k) \leqslant W^*, \tag{7}$$

donde θ_k son puntos de cierto conjunto numerable $T = \{\theta_1, \theta_2, \ldots\}$ siempre denso en D. Según la definición del precio superior de W^*, existe una sucesión de estrategias $\pi_n = \pi_n(x, \cdot)$ tal, que

$$W(\pi_n, \theta_k) < W^* + 1/n \tag{8}$$

para todos k.

Ahora, mediante las distribuciones π_n construyamos la sucesión de elementos aleatorios especialmente seleccionados ζ_n y separemos de ella la subsucesión convergente. Para esto, designemos por $f_{\theta_k}(x)$ la densidad de la distribución $\mathbf{P}_{\theta_k}$ respecto a la medida probabilística $\mu = \sum_{j=1}^{\infty} 2^{-k}\mathbf{P}_{\theta_k}$, así que

$$W(\pi_n, \theta_k) = \int\int w(u, \theta_k)\pi_n(x, du) f_{\theta_k}(x)\mu(dx).$$

Examinemos el espacio $D \times R^T$, donde R^T es el espacio de los valores de los elementos $f(x) = \{f_{\theta_1}(x), f_{\theta_2}(x), \ldots\}$ con σ-álgebra $\mathfrak{B}^T$ engendrada por los conjuntos cilíndricos. Pongamos a cada estrategia π en correspondencia con el espacio probabilístico $(D \times \mathscr{X}, \sigma_D \mathfrak{B}_{\mathscr{X}}, \mathbf{P})$, donde la distribución $\mathbf{P}$ es definida por la igualdad

$$\mathbf{P}(\delta \in A, X \in B) = \int_B \mu(dx)\pi(x, A), \quad A \in \sigma_D, \quad A \in \mathfrak{B}_X. \tag{10}$$

Definamos en este espacio los elementos aleatorios $\zeta = \zeta(\delta; X) = (\delta; f_{\theta_1}(X), f_{\theta_2}(X), \ldots) = (\delta; f(X))$ y designamos por ζ_n los elementos correspondientes a π_G, así que ζ_n son variables aleatorias en el espacio probabilístico muestral $(D \times R^T, \sigma_D \times \mathfrak{B}^T, \mathbf{\Pi}_n)$, y la distribución $\mathbf{\Pi}_n$ ha sido engendrada por π_n, por la fórmula (10) y por la aplicación $\zeta(\delta, x)$: $D \times \mathscr{X} \to D \times R^T$.

Designemos por $\mathbf{\Pi}_n^{(k)}$ las contracciones de la distribución $\mathbf{\Pi}_n$ en $D \times R^k$ (es la distribución compatible $(\delta; f_{\theta_1}(X), \ldots, f_{\theta_k}(X))$, y por λ la distribución $f(X)$ en $(\mathscr{X}, \mathfrak{B}_{\mathscr{X}}, \mu)$. Necesitaremos el

Lema 4. *Existe tal disitribución $\overline{\mathbf{\Pi}}$ en el espacio medible $(D \times R^T, \sigma_D \times \mathfrak{B}^T)$ y tal subsucesión $\{\pi_{n'}\} \subset \{\pi_n\}$, que*

$$\mathbf{\Pi}_{n'}^{(k)} \to \overline{\mathbf{\Pi}}^{(k)} \tag{11}$$

para cualquier k ($\overline{\Pi}^{(k)}$ *son las contracciones de* $\overline{\Pi}$),

$$\overline{\Pi}(D \times C) = \lambda(C), \quad C \in \mathfrak{B}^T. \tag{12}$$

La demostración del lema 4 se ofrecerá más tarde.

Designemos por $\bar{\zeta} = (\bar{\delta}; \bar{f})$ cierto elemento aleatorio con distribución $\overline{\Pi}$. La relación (12) significa que la distribución $\bar{f}$ coincide con λ (la segunda "coordenada" ζ_n no modifica la distribución al variar n). Como el espacio D constituye un compacto métrico, el mismo es separable y, por consiguiente, (véase [38], p. 191) existe cierta distribución condicional (regular) $\bar{\delta}$ respecto a $\bar{f}(X)$, la cual designaremos por $\Pi(\cdot/\bar{f}(x))$.

Examinemos la estrategia $\bar{\pi}(x, A) = \Pi(\bar{\delta} \in A/\bar{f}(X))$ y demostremos que para ella se cumple (7).

Señalemos previamente que

$$\mathbf{M}w(\bar{\delta}, \theta_k)\bar{f}_{\theta_k} = \mathbf{M}f_{\theta_k}\mathbf{M}(w(\bar{\delta}, \theta_k)/X) = \int f_{\theta_k}(x)\int w(u, \theta_k)\bar{\pi}(u, dx)\mu(dx) = W(\bar{\pi}, \theta_k). \tag{13}$$

Seguidamente, en virtud del lema 4, la distribución $(\delta_{n^*}, f_{\theta_k}(X))$ converge débilmente hacia la distribución $(\bar{\delta}, \bar{f}_{\theta_k}(X))$. Como la función w es continua, la distribución compatible $(w(\delta_{n^*}, \theta_k), f_{\theta_k}(X))$ converge débilmente hacia la distribución $(w(\bar{\delta}, \theta_k), \bar{f}_{\theta_k}(X))$. Pero la función $g(u, v) = w(u, \theta_k)v$ es continua respecto a u y v y es mayorada por la función $g(v) = cv$, $c = \max_u w(u, \theta_k)$ tal, que $\mathbf{M}g(f_{\theta_k}(X)) = c\int f_{\theta_k}(x)\mu(dx) = c$. Por eso, según el teorema de continuidad para los momentos (véase el teorema 1.5.4),

$$\lim_{n^* \to \infty} \mathbf{M}g(\delta_{n^*}, f_{\theta_k}(X)) = \mathbf{M}g(\bar{\delta}, \bar{f}_{\theta_k}(X)),$$

o bien, que es lo mismo, $\lim\limits_{n^* \to \infty} \mathbf{M}w(\delta_{n^*}, \theta_k)f_{\theta_k}(X) = \mathbf{M}w(\bar{\delta}, \theta_k)\bar{f}_{\theta_k}(X)$.

En virtud de (9) y (13), esto nos ofrece la convergencia

$$\lim_{n^* \to \infty} W(\pi_n, \theta_k) = W(\bar{\pi}, \theta_k).$$

En vista de que el primer miembro de esta igualdad (véase (8)) no supera W^*, el lema 3 queda demostrado.

Demostración del lema 4. Fijemos cualquier $k \geqslant 1$ y examinemos $D \times R^k$ como espacio separable métrico completo respecto a la métrica engendrada por la métrica euclídea en R^k y la métrica ϱ_D. Para cualquier $\varepsilon > 0$ en R^k habrá un compacto K_ε tal, que $\mathbf{P}((f_{\theta_1}(X), \ldots, f_{\theta_k}(X)) \in K_\varepsilon) \geqslant 1 - \varepsilon$. $D \times K_\varepsilon$ es un compacto un $D \times R^k$ y como

$$\mathbf{P}(\delta_n \in D, \ \ (f_{\theta_1}(X), \ \ldots, f_{\theta_k}(X)) \in K_\varepsilon) \geqslant 1 - \varepsilon,$$

la sucesión de las distribuciones $\Pi_n^{(k)}$ es densa (véase [5]). Por consiguiente, según el teorema de Prójorov [5], existe una distribución $\overline{\Pi}^{(k)}$ y una subsucesión $n^{(k)} = (n_1^{(k)}, n_2^{(k)}, \ldots)$ tales, que $\Pi_{n^{(k)}}^{(k)} \Rightarrow \overline{\Pi}^{(k)}$. Pero las distribuciones $\overline{\Pi}^{(k)}$, evidentemente, se hallan en concordancia y, por consiguiente, según el teorema de Kolmogórov, en $(D \times R^T, \sigma_D \times \mathfrak{B}^T)$ existe cierta distribución $\overline{\Pi}$ para la cual $\overline{\Pi}^{(k)}$ son las contracciones en $(D \times R^k, \sigma_D \times \mathfrak{B}^k)$.

Por otro lado, podemos considerar que $n^{(k+1)} \subset n^{(k)}$. Poniendo $n^* = (n_1^{(1)}, n_2^{(2)}, n_3^{(3)}, \ldots)$ obtendremos una subsucesión para la cual $\Pi_{n^*}^{(k)} \Rightarrow \overline{\Pi}^{(k)}$ con todos los valores de k.

Demostremos ahora (12). Sea $C \in \mathfrak{B}^T$ un conjunto cilíndrico tal, que la $\overline{\Pi}$-medida de su frontera es igual a cero. Designemos por $C^{(k)} = C \cap R^k \in \mathfrak{B}^k$ el conjunto de R^k formado por las primeras k coordenadas de los puntos de C, y pongamos $\overline{C}^{(k)} = C^{(k)} \times R^{T-k} \in \mathfrak{B}^T$.

Entonces $\lambda(\overline{C}^{(k)}) = \Pi_{n^*}^{(k)}(D \times C^{(k)}) \to \overline{\Pi}^{(k)}(D \times C^{(k)})$. Como $\overline{C}^{(k+1)} \subset \overline{C}^{(k)}$, $C = \bigcap\limits_{k=1}^{\infty} \overline{C}^{(k)}$, entonces

$$\lambda(C) = \lim_{k \to \infty} \lambda(\overline{C}^{(k)}) = \lim_{k \to \infty} \overline{\Pi}^{(k)}(D \times C^{(k)}) = \lim_{k \to \infty} \overline{\Pi}(D \times \overline{C}^{(k)}) = \overline{\Pi}(D \times C).$$

El lema 4 queda demostrado.

Tabla I. Distribución normal $\Phi_{0,1}$

En la tabla se dan los valores de

$$\overline{\Phi}(x) = \Phi_{0,1}(x, \infty)) = \frac{1}{\sqrt{2\pi}}\int_{x}^{\infty} e^{-t^2/2}dt.$$

Tabla I

x	0	1	2	3	4
0,0	0,5000	0,4960	0,4920	0,4880	0,4840
0,1	,4602	,4562	,4522	,4483	,4443
0,2	,4207	,4168	,4129	,4090	,4052
0,3	,3821	,3783	,3745	,3707	,3669
0,4	,3446	,3409	,3372	,3336	,3300
0,5	,3085	,3050	,3015	,2981	,2946
0,6	,2743	,2709	,2676	,2643	,2611
0,7	,2420	,2389	,2358	,2327	,2297
0,8	,2119	,2090	,2061	,2033	,2005
0,9	,1841	,1814	,1788	,1762	,1736
1,0	,1587	,1562	,1539	,1515	,1492
1,1	,1357	,1335	,1314	,1292	,1271
1,2	,1151	,1131	,1112	,1093	;1075
1,3	,0968	,0951	,0934	,0918	,0901
1,4	,0808	,0793	,0778	,0764	,0749
1,5	,0668	,0655	,0643	,0630	,0618
1,6	,0548	,0537	,0526	,0516	,0505
1,7	,0446	,0436	,0427	,0418	,0409
1,8	,0359	,0351	,0344	,0336	,329
1,9	,0288	,0281	,0274	,0268	,0262
2,0	,0228	,0222	,0217	,0212	,0207
2,1	,0179	,0174	,0170	,0166	,0162
2,2	,0139	,0136	,0132	,0129	,0125
2,3	,0107	,0104	,0102	,0099	,0096
2,4	,0082	,0080	,0078	,0075	,0073
2,5	,0062	,0060	,0059	,0057	,0055
2,6	,0047	,0045	,0044	,0043	,0041
2,7	,0035	,0034	,0033	,0032	,0031
2,8	,0026	,0025	,0024	,0023	,0023
2,9	,0019	,0018	,0018	,0017	,0016
$x =$	3,0	3,1	3,2	3,3	3,4
$\overline{\Phi}(x) =$	0,0013	0,0010	0,0007	0,0005	0,0003

Tabla I (continuación)

x	5	6	7	8	9
0,0	0,4810	0,4761	0,4721	0,4681	0,4641
0,1	,4404	,4364	,4325	,4286	,4247
0,2	,4013	,3974	,3936	,3897	,3859
0,3	,3632	,3594	,3557	,3520	,3483
0,4	,3264	,3228	,3192	,3156	,3121
0,5	,2912	,2877	,2843	,2810	,2776
0,6	,2578	,2546	,2514	,2483	,2451
0,7	,2266	,2236	,2206	,2177	,2148
0,8	,1977	,1949	,1922	,1894	,1867
0,9	,1711	,1685	,1660	,1635	,1611
1,0	,1469	,1446	,1423	,1401	,1379
1,1	,1251	,1230	,1210	,1190	,1170
1,2	,1056	,1038	,1020	,1003	,0985
1,3	,0885	,0869	,0853	,0838	,0823
1,4	,0735	,0721	,0708	,0694	,0681
1,5	,0606	,0594	,0582	,0571	,0559
1,6	,0495	,0485	,0475	,0465	,0455
1,7	,0401	,0392	,0384	,0375	,0367
1,8	,0322	,0314	,0307	,0301	,0294
1,9	,0256	,0250	,0244	,0239	,0233
2,0	,0202	,0197	,0192	,0188	,0183
2,1	,0158	,0154	,0150	,0146	,0143
2,2	,0122	,0119	,0116	,0113	,0110
2,3	,0094	,0091	,0089	,0087	,0084
2,4	,0071	,0069	,0068	,0066	,0064
2,5	,0054	,0052	,0051	,0049	,0048
2,6	,0040	,0039	,0038	,0037	,0036
2,7	,0030	,0029	,0028	,0027	,0026
2,8	,0022	,0021	,0021	,0020	,0019
2,9	,0016	,0015	,0015	,0014	,0014
$x =$	3,5	3,6	3,7	3,8	3,9
$\overline{\Phi}(x) =$	0,0002	0,0002	0,0001	0,0001	0,0000

Tabla II. Cuantilas de la distribución normal

En la tabla se dan los valores de λ_ε tales, que

$$\overline{\Phi}(\lambda_\varepsilon) = \Phi_{0,1}((\lambda_\varepsilon, \infty)) = \varepsilon.$$

Tabla II

100_ε	λ_ε	100_ε	λ_ε	100_ε	λ_ε
50	0,0000	20	0,8416	0,5	2,5758
45	0,1257	15	1,0364	0,1	3,0902
40	0,2533	10	1,2816	0,05	3,2905
35	0,3853	5	1,6449	0,01	3,7190
30	0,5244	2,5	1,9600	0,005	3,8906
25	0,6745	1	2,3263		

Tabla III. Distribución ji-cuadrado $\mathbf{H}_k$

En la tabla se dan los valores (véanse el §§ 2.2)

$$H_k(x) = \mathbf{H}_k((x, \infty)) = \frac{1}{2^{k/2}\Gamma(k/2)} \int_x^\infty t^{k/2-1} e^{-t/2} dt$$

cuando $1 \leqslant k \leqslant 20$. Para mayores valores de k se puede utilizar la aproximación (véase el § 2.2, tabla I)

$$H_k(x) \approx \overline{\Phi}(\sqrt{2x} - \sqrt{2k-1}) \equiv \hat{H}_k(x). \tag{1}$$

La última columna de la tabla contiene los valores de $\hat{H}_k(x)$ cuando $k = 20$. Comparándolos con los valores dados en la columna anterior se puede estimar el grado de precisión de la aproximación (1). Con el aumento de k disminuye el error.

Tabla III

x \ k	1	2	3	4	5
0,1	0,7518	0,9512	0,9918	0,9988	0,9998
,2	,6547	,9048	,9776	,9953	,9991
,4	,5271	,8187	,9402	,9825	,9953
,6	,4386	,7408	,8964	,9631	,9880
,8	,3711	,6703	,8495	,9385	,9770
1,0	,3173	,6065	,8013	,9098	,9626
,5	,2207	,4724	,6823	,8266	,9131
2	,1573	,3679	,5725	,7358	,8492
3	,0833	,2231	,3916	,5578	,7000
4	,0455	,1353	,2615	,4060	,5494
5	,0254	,0821	,1718	,2873	,4159
6	,0143	,0498	,1116	,1992	,3062
7	,0082	,0302	,0719	,1359	,2206
8	,0047	,0183	,0460	,0916	,1562
9	,0027	,0111	,0293	,0611	,1091
10	,0016	,0067	,0186	,0404	,0752
11	,0009	,0041	,0117	,0266	,0514
12	,0005	,0025	,0074	,0174	,0348
13	,0003	,0015	,0046	,0113	,0234
14	,0002	,0009	,0029	,0073	,0156
15	,0001	,0006	,0018	,0047	,0104
16	,0001	,0003	,0011	,0030	,0068
17		,0002	,0007	,0019	,0045
18		,0001	,0004	,0012	,0030
19		,0001	,0003	,0008	,0019
20		,0001	,0002	,0001	,0013
21			,0001	,0003	,0008
22			,0001	,0002	,0005
23				,0001	,0003
24				,0001	,0002
25				,0001	,0001

Tabla III (continuación)

x \ k	6	7	8	9	10
0,5	0,9978	0,9995	0,9999	1,0000	1,0000
1,0	,9856	,9948	,9983	0,9994	0,9998
,5	,9595	,9823	,9927	,9972	,9989
2,0	,9197	,9598	,9810	,9915	,9963
,5	,8685	,9271	,9617	,9809	,9909
3	,8089	,8850	,9344	,9643	,9814
4	,6767	,7798	,8571	,9114	,9474
5	,5438	,6600	,7576	,8343	,8912
6	,4232	,5398	,6472	,7399	,8153
7	,3204	,4284	,5366	,6371	,7254
8	,2381	,3326	,4335	,5342	,6288
9	,1736	,2527	,3423	,4373	,5321
10	,1246	,1886	,2650	,3505	,4405
11	,0884	,1386	,2017	,2757	,3575
12	,0620	,1006	,1512	,2133	,2851
13	,0430	,0721	,1119	,1626	,2237
14	,0296	,0512	,0818	,1223	,1730
15	,0203	,0360	,0592	,0909	,1321
16	,0138	,0251	,0424	,0669	,0996
17	,0093	,0174	,0301	,0487	,0744
18	,0062	,0120	,0212	,0352	,0550
19	,0042	,0082	,0149	,0252	,0403
20	,0028	,0056	,0103	,0179	,0193
21	,0018	,0038	,0072	,0127	,0211
22	,0012	,0025	,0049	,0084	,0151
23	,0008	,0017	,0034	,0062	,0108
24	,0005	,0011	,0023	,0043	,0076
25	,0003	,0008	,0016	,0030	,0054
26	,0002	,0005	,0011	,0020	,0037
27	,0002	,0003	,0007	,0014	,0026
28	,0001	,0002	,0004	,0010	,0018
29	,0001	,0002	,0003	,0007	,0013
30		,0001	,0002	,0004	,0009

Tabla III (continuación)

x \ k	11	12	13	14	15
2	0,9985	0,9994	0,9998	0,9999	1,0000
3	,9907	,9955	,9979	,9991	0,9996
4	,9699	,9834	,9912	,9955	,9977
5	,9312	,9580	,9752	,9858	,9921
6	,8734	,9161	,9462	,9665	,9798
7	,7991	,8576	,9022	,9347	,9577
8	,7133	,7852	,8436	,8893	,9238
9	,6219	,7029	,7729	,8311	,8775
10	,5304	,6160	,6939	,7622	,8197
12	,3636	,4457	,5276	,6063	,6790
14	,2330	,3007	,3738	,4497	,5255
16	,1411	,1912	,2491	,3134	,3821
18	,0816	,1157	,1575	,2068	,2627
20	,0453	,0671	,0952	,1301	,1719
21	,0334	,0504	,0729	,1016	,1368
22	,0244	,0375	,0554	,0786	,1078
23	,0177	,0277	,0417	,0603	,0841
24	,0127	,0203	,0311	,0458	,0651
25	,0091	,0148	,0231	,0346	,0499
26	,0065	,0107	,0170	,0259	,0380
27	,0046	,0077	,0124	,0193	,0287
28	,0032	,0055	,0091	,0142	,0216
29	,0023	,0039	,0066	,0105	,0161
30	,0016	,0028	,0047	,0076	,0119
31	,0011	,0020	,0034	,0055	,0088
32	,0008	,0014	,0024	,0040	,0064
33	,0005	,0010	,0017	,0029	,0047
34	,0004	,0007	,0012	,0021	,0034
35	,0003	,0005	,0009	,0015	,0025
36	,0002	,0003	,0006	,0010	,0018
37	,0001	,0002	,0004	,0007	,0013
38	,0001	,0002	,0003	,0005	,0009
39	,0001	,0001	,0002	,0004	,0006
40		,0001	,0001	,0003	,0005

Tabla III (continuación)

x \ k	16	17	18	19	20	$P_{20}(x)$
4	0,9989	0,9995	0,9998	0,9999	1,000	0,9997
5	,9958	,9978	,9989	,9994	,0,9997	,9990
6	,9881	,9932	,9962	,9979	,9989	,9973
7	,9733	,9836	,9901	,9942	,9967	,9938
8	,9489	,9666	,9786	,9867	,9919	,9876
9	,9134	,9403	,9597	,9735	,9829	,9774
10	,8666	,9036	,9319	,9530	,9682	,9619
12	,7440	,8001	,8472	,8856	,9161	,9109
14	,5987	,6671	,7291	,7837	,8305	,8298
16	,4530	,5238	,5926	,6573	,7166	,7218
18	,3239	,3888	,4557	,5224	,5874	,5968
20	,2202	,2742	,3328	,3946	,4579	,4683
22	,1432	,1847	,2320	,2843	,3405	,3489
24	,0895	,1194	,1550	,1962	,2424	,2472
26	,0540	,0745	,0998	,1302	,1658	,1670
28	,0316	,0449	,0621	,0834	,1094	,1078
30	,180	,0264	,0375	,0518	,0699	,0667
31	,0135	,0200	,0288	,0404	,0552	,0517
32	,0100	,0151	,0220	,0313	,0433	,0396
33	,0074	,0113	,0167	,0240	,0337	,0301
34	,0054	,0084	,0126	,0184	,0261	,0227
35	,0040	,0062	,0095	,0140	,0201	,0169
36	,0029	,0046	,0071	,0106	,0154	,0125
37	,0021	,0034	,0052	,0080	,0117	,0092
38	,0015	,0025	,0039	,0059	,0089	,0067
39	,0011	,0018	,0029	,0044	,0067	,0048
40	,0008	,0013	,0021	,0033	,0050	,0035
41	,0006	,0009	,0015	,0024	,0037	,0025
42	,0004	,0007	,0011	,0018	,0028	,0017
43	,0003	,0005	,0008	,0013	,0020	,0012
44	,0002	,0003	,0006	,0009	,0015	,0010
45	,0001	,0002	,0004	,0007	,0011	,0006

Tabla IV. Distribución de Student T_k

En la tabla se dan los valores de

$$T_k(x) = T_k((x, \infty)) = \frac{\Gamma(k+1)/2}{\sqrt{k\pi}\,\Gamma(k/2)} \int_x^\infty (1 + t^2/k)^{-(k+1)/2} dt$$

cuando $1 \leqslant x \leqslant 20$. Para mayores valores de k se puede utilizar la aproximación (véase el § 2.2, tabla I)

$$T_k(x) \approx \bar{\Phi}(x) = \Phi_{0,1}((x, \infty)). \tag{2}$$

La exactitud de aproximación (2) cuando $k = 20$ se puede apreciar comparando la última columna de la tabla con la tabla I.

Tabla IV

x \ k	1	2	3	4	5
0,0	0,5000	0,5000	0,5000	0,5000	0,5000
5	,3524	,3333	,3257	,3217	,3191
1,0	,2500	,2113	,1955	,1869	,1816
2	,2211	,1765	,1581	,1482	,1419
4	,1974	,1482	,1280	,1170	,1102
6	,1778	,1253	,1039	,0924	,0852
8	,1614	,1068	,0848	,0731	,0659
2,0	,1476	,0917	,0697	,0581	,0510
2	,1358	,0794	,0576	,0463	,0395
4	,1257	,0692	,0479	,0372	,0308
6	,1169	,0679	,0402	,0300	,0241
8	,1092	,0537	,0339	,0244	,0190
3,0	,1024	,0477	,0282	,0200	,0150
2	,0964	,0427	,0247	,0165	,0120
4	,0910	,0383	,0212	,0136	,0096
6	,0862	,0346	,0184	,0114	,0078
8	,0819	,0314	,0160	,0095	,0063
4,0	,0780	,0286	,0140	,0081	,0052
2	,0744	,0261	,0123	,0068	,0045
4	,0711	,0240	,0109	,0058	,0035
6	,0681	,0221	,0097	,0050	,0029
8	,0654	,0204	,0086	,0043	,0024
5,0	,0628	,0199	,0077	,0037	,0020
2	,0605	,0175	,0069	,0033	,0017
4	,0583	,0163	,0062	,0028	,0015
6	,0562	,0152	,0056	,0025	,0012

Tabla IV (continuación)

x \ k	1	2	3	4	5
8	,0543	,0142	,0051	,0022	,0011
6,0	,0526	,0133	,0046	,0019	,0009
2	,0509	,0125	,0042	,0017	,0008
4	,0493	,0118	,0039	,0015	,0007
6	,0479	,0111	,0035	,0014	,0006
8	,0465	,0105	,0033	,0012	,0005
7,0	,0452	,0099	,0030	,0011	,0005
2	,0439	,0094	,0028	,0010	,0004
4	,0428	,0089	,0025	,0009	,0004
6	,0416	,0086	,0024	,0008	,0003
8	,0406	,0080	,0022	,0007	,0003
8,0	,0396	,0076	,0020	,0007	,0002

Tabla IV (continuación)

x \ k	6	7	8	9	10
0,0	0,5000	0,5000	0,5000	0,5000	0,5000
5	,3174	,3162	,3153	,3145	,3139
1,0	,1780	,1753	,1733	,1717	,1704
2	,1377	,1346	,1322	,1304	,1280
4	,1055	,1021	,0995	,0975	,0959
6	,0804	,0768	,0741	,0720	,0703
8	,0610	,0574	,0548	,0527	,0510
2,0	,0462	,0428	,0403	,0383	,0367
2	,0350	,0319	,0295	,0277	,0262
4	,0266	,0237	,0216	,0199	,0186
6	,0203	,0177	,0158	,0144	,0132
8	,0156	,0132	,0116	,0104	,0094
3,0	,0120	,0100	,0085	,0075	,0067
2	,0093	,0075	,0063	,0054	,0047
4	,0072	,0057	,0047	,0039	,0034
6	,0057	,0044	,0035	,0029	,0024
8	,0045	,0034	,0026	,0022	,0017
4,0	,0035	,0026	,0020	,0015	,0013
2	,0028	,0020	,0015	,0012	,0009
4	,0023	,0016	,0011	,0009	,0007
6	,0018	,0012	0,009	,0006	,0005
8	,0015	,0010	,0007	,0005	,0004
5,0	,0012	,0008	,0005	,0004	,0003

Tabla IV (continuación)

x \ k	11	12	13	14	15
0,0	0,5000	0,5000	0,5000	0,5000	0,5000
5	,3135	,3131	,3127	,3124	,3112
1,0	,1694	,1685	,1678	,1671	,1666
2	,1277	,1266	,1258	,1250	,1244
4	,0945	,0934	,0925	,0916	,0909
6	,0689	,0678	,0668	,0660	,0652
8	,0496	,0485	,0475	,0467	,0460
2,0	,0354	,0343	,0334	,0326	,0320
2	,0250	,0241	,0232	,0225	,0219
4	,0176	,0168	,0160	,0154	,0149
6	,0123	,0116	,0110	,0105	,0100
8	,0086	,0080	,0075	,0071	,0067
3,0	,0060	,0055	,0051	,0048	,0045
2	,0042	,0038	,0035	,0032	,0030
4	,0030	,0026	,0024	,0022	,0020
6	,0021	,0018	,0016	,0014	,0013
8	,0015	,0013	,0011	,0010	,0009
4,0	,0010	,0009	,0008	,0007	,0006

Tabla IV (continuación)

x \ k	16	17	18	19	20
0,0	0,5000	0,5000	0,5000	0,5000	0,5000
5	,3119	,3117	,3116	,3114	,3113
1,0	,1661	,1657	,1653	,1649	,1646
2	,1238	,1233	,1228	,1224	,1221
4	,0903	,0898	,0893	,0888	,0884
6	,0646	,0640	,0635	,0630	,0626
8	,0454	,0448	,0443	,0439	,0435
2,0	,0314	,0309	,0304	,0300	,0296
2	,0214	,0210	,0205	,0202	,0199
4	,0145	,0141	,0137	,0134	,0131
6	,0097	,0093	,0090	,0082	,0086
8	,0064	,0061	,0059	,0057	,0055
3,0	,0042	,0040	,0038	,0037	,0035
2	,0028	,0026	,0025	,0024	,0022
4	,0018	,0017	,0016	,0015	,0014
6	,0012	,0011	,0010	,0009	,0009
8	,0008	,0007	,0007	,0006	,0006
4,0	,0005	,0005	,0004	,0004	,0003

Observaciones bibliográficas

Más abajo se aducen algunos comentarios bibliográficos en los que se hacen intentos de seguir la historia de aparición de las ideas y los resultados fundamentales expuestos en este libro. Dichos comentarios no pretenden ser completos y a menudo contendrán referencias no a artículos originales poco abordables, sino a manuales, monografías o artículos de resumen, en los que es más fácil hallar los resultados necesarios. Por ejemplo, en [95] y [57] se ofrecen indicaciones bibliográficas e informaciones históricas más amplias.

Algunos conceptos fundamentales de la estadística matemática surgieron ya a principios del siglo pasado y están relacionados con los nombres de Laplace y Gauss. A finales del siglo pasado, los trabajos de K. Pearson dieron comienzo a un período de desarrollo intenso de dicha ciencia. El mismo ha sido condicionado por las obras fundamentales de R. Fisher, J. Neyman, A. N. Kolmogórov y A. Wald. En la Unión Soviética, el desarrollo de la estadística matemática se halla relacionado, antes que nada, con los nombres de A. N. Kolmogórov y N. V. Smirnov.

Capítulo 1

§§ 2—4. El teorema de Glivenko — Cantelli fue establecido en el año (a Glivenko le pertenece su demostración para una distribución continua, y a Cantelli, para el caso general).

La demostración del teorema 1.2.2 se asemeja a la expuesta en [61], p. 28, y es un caso particular de utilización de un enfoque más general basado en la "aproximación finita" de la clase de conjuntos sujetos a estudio. En su forma completa, este enfoque se ofrece en el Suplemento I, donde ha sido demostrado el teorema 1.4.2. Un enfoque análogo fue examinado independientemente en [27]. La ley del logaritmo reiterado (teorema 1.4.3) fue establecida en [52].

§ 6. Los teoremas 1.6.1 y 1.6.2 de la distribución de $nF_n^*(t)$ se dan en el libro de Feller [32], t. 2, § 3, cap. III. El teorema 1.6.3 de la convergencia del proceso $\sqrt{n}(F_n^*(t) - F(t))$ hacia el puente browniano, demostrado en el Suplemento II, fue establecido por Donsker en [28]. Una demostración algo diferente (en comparación con el Suplemento II) del teorema 1.6.3 se ofrece en la obra de Billingsley [5].

§ 7. La afirmación del ejemplo 1.7.3 acerca de la distribución límite de la estadística $\chi^2(X)$ (ji-cuadrado) fue por primera vez obtenida por K. Pearson (véase [25], p. 454).

§ 8. La afirmación del corolario 1.8.2 constituye el contenido del teorema de Kolmogórov, y la del corolario 1.8.3, el del teorema de Smirnov. Este último también comprende la forma explícita de la distribución de $\int_0^1 [w^0(t)]^2 dt$, que omitimos debido a su complejidad (véase [78]).

§ 10. Las estimaciones de la densidad que se examinan en este párrafo fueron introducidas por Parzen [72] y Rosenblatt [79]. La bibliografía y el análisis de los resultados en esta dirección se exponen en el trabajo de resumen de Rosenblatt [80] y en el § 25 del libro de Chentsov [19].

Capítulo 2

§ 2. Algunas otras familias paramétricas se describen en el libro de Wilks [93]. Una investigación muy completa de las distribuciones de los términos de la serie variacional fue llevada a efecto B. V. Gnedenko. Una exposición completa de los resultados y una amplia bibliografía al respecto se pueden hallar en la obra de David [26].

§ 4. El método de momentos es, históricamente, el primer método regular de construcción de las estimaciones. El mismo fue propuesto por K. Pearson en 1894.

§ 5. El método del mínimo χ^2 fue propuesto por R. Fisher en 1922.

§ 6. El método de verosimilitud máxima en casos particulares fue empleado aún por Gauss. Como método general para obtener las estimaciones, el mismo fue propuesto por Fisher en 1912 en un artículo breve. Más tarde, en 1925, Fisher estudió las propiedades asintóticas de la e.v.m. en su obra clásica [35].

§§ 7 y 8. Los enfoques expuestos, dedicados a la comparación de las estimaciones, son universalmente reconocidos. Hemos adoptado la demostración del lema 2.7.3 dada en [25]. El concepto de estimación eficiente fue introducido en 1922 por Fisher en [34].

§§ 9 y 10. El concepto fundamental de esperanza matemática condicional fue introducido en 1933 por A. N. Kolmogórov en su obra clásica [54]. Las propiedades de las distribuciones condicionales fueron detalladamente estudiadas en [38], [30] y [84].

§ 11. El enfoque bayesiano ha sido ampliamente utilizado por Laplace aún en el siglo pasado. Este enfoque fue criticado por Fisher, y en los años 20 y 30 de nuestro siglo, el centro de gravedad de las investigaciones se desplazó hacia las estimaciones eficientes y asintóticamente eficientes. Más tarde, a medida que se concebía el papel fundamental del enfoque bayesiano, otra vez comenzó a crecer el interés por este último.

El concepto de estimación minimax se introdujo en la estadística matemática junto con el enfoque de la teoría de los juegos, desarrollado en los trabajos de Borel (1921) y J. Neyman (1928); los teoremas 2.11.1—2.11.3 fueron obtenidos por Hodges y Lehman [44].

§ 12. El concepto fundamental de la estadística suficiente fue introducido en 1922 por R. Fisher en [34], quien, y más tarde J. Neyman [66], propusieron un criterio simple que revela la existencia y el tipo de estadística suficiente. Este criterio lleva el nombre de teorema de factorización de Neyman — Fisher y está representado en el teorema 2.12.1. La estricta demostración del teorema de Neyman — Fisher, desde el punto de vista de la teoría de los conjuntos, fue obtenida tan sólo en 1949 por Halmos y Savage [43].

§ 13. El concepto de σ-álgebra suficiente es más amplio que el concepto de estadística suficiente. Las condiciones necesarias y suficientes para su coincidencia se dan en [95]. Tanto la construcción de las particiones suficientes como el teorema 2.13.1 están relacionados con el trabajo de Lehmann y Scheffe [59] dedicado a la aclaración de las condiciones de existencia y a la construcción de las estadísticas mínimas suficientes. La exposición breve de este artículo se ofrece en [95]. La demostración del teorema 2.13.2 le pertenece a I. S. Borísov.

§ 14. El teorema 2.14.1 fue independientemente obtenido por Blackwell [6] (1947), Rao [75] (1945), [76] (1949) y Kolmogórov [53] (1950). Los autores del teorema 2.14.3 son Rao [76] (1949) y Blackwell [6] (1947).

§ 15. La familia exponencial ha sido mencionada por Fisher aún en [34], pero su importancia teórica fue concebida en los años 30 en las obras de Pitman, Kupman y Darmois. Por eso dicha familia a veces lleva los nombres de estos científicos. El teorema 2.15.2 fue demostrado por Lehmann ([57], p. 183).

§§ 16 y 17. La desigualdad de Rao — Cramer a veces también se denomina desigualdad de información. De hecho, ésta pertenece a Fisher [35], aunque en la forma expuesta fue independientemente obtenida por Frechet [37] en 1943, Rao [74] en 1945 y Cramer [24] en 1946.

Las condiciones de regularidad, necesarias para el cumplimiento de la desigualdad, en los manuales de estadística matemática no siempre se interpretan correctamente. Se trata de las condiciones que aseguran la validez de la derivación respecto al parámetro bajo el signo integral. La demostración de dicha validez a menudo contiene lagunas (véase por ejemplo, [95]) o su exposición no se ofrece en absoluto (por ejemplo, en [86]). En una serie de casos, la misma se menciona en forma de condición [86]), lo cual no es cómodo para la verificación en problemas reales.

Las condiciones de regularidad adoptadas en el libro son muy simples, aunque, por lo visto, no son las más generales (compárense con ([48]). El hecho de que en estas condiciones se pueda derivar bajo el signo integral, fue demostrado en el Suplemento VI escrito a base de los resultados obtenidos por A. I. Sajanenko.

En [95] y [19] se ofrecen distintas generalizaciones de la desigualdad de Rao — Cramer. El concepto de información (de Fisher) fue introducido en [35]. Al demostrar los teoremas 2.16.1A y 2.17.1 nos hemos guiado por los libros [95] y [48].

§§ 18 y 19. A Hotelling y Pítman les pertenece la idea de utilizar las consideraciones invariantes. S. Stein contribuyó considerablemente al desarrollo de la teoría. El contenido principal del teorema 2.18.1 le pertenece a Pitman. Al demostrarlo hemos utilizado las exposiciones en [95] y [48]. El carácter minimax de la estimación de Pitman fue establecido por Girchik y Savage.

§ 20. Los resultados de este párrafo fueron obtenidos por el autor junto con A. I. Sajanenko [13]. Cuando las limitaciones son más rígidas, algunas desigualdades también se pueden obtener de las obras [40] y [18].

§ 21. En el caso paramétrico, la distancia de Kullback — Leibler también se llama función de información de Kullback — Leibler. Al describir las probabilidades de las grandes divergencias de la distribución empírica, I. N. Sanov llegó independientemente a la referida distancia. La idea del amplio uso de la distancia de Hellinger para estudiar las propiedades de la relación de verosimilitud fue adoptada del libro de Ibraguímov y Jasminski [48]. Las demostraciones de los principales teoremas del § 23 también se basan en los resultados de este libro. La demostración del teorema 2.21.3 ha sido considerablemente simplificada por A. I. Sajanenko.

§ 22. El teorema 2.22.1 fue establecido en 1952 por Chapman y Robbins en [17] y en 1952 por Kiefer en [51].

§§ 23—25. Se expone el material de nuestras conferencias, perfeccionado considerablemente después de la aparición del libro de Ibraguímov y Jasminski [48]. Los prefeccionamientos principales están relacionados con la utilización sistemática de la distancia de Hellinger para estimar $\mathbf{M}_\theta Z^{1/2}(u)$. A. I. Sajanenko propuso utilizar $\int \mathbf{M}_\theta |(Z^{3/4}(u))| \, du$ para estimar $\sup_u Z(u)$ (véanse los teoremas 2.23.1 y 2.23.2). Aún Fisher, en [35], estableció la normalidad asintótica y la eficacia asintótica de la e.v.m. Condiciones muy generales de la normalidad asintótica de la e.v.m. fueron obtenidas en [48].

La normalidad asintótica de la densidad a posteriori (o de la relación de verosimilitud) fue descubierta por S. N. Bernshtein en 1927. El teorema 2.25.4 pertenece a Bahadur [1]. Los caracteres asintóticamente bayesiano y asintóticamente minimax de la e.v.m. se obtienen fácilmente merced a los resultados del § 2.20. Antes, el carácter asintóticamente bayesiano de la e.v.m. se establecía con limitaciones más rígidas para la densidad de la distribución a priori.

Hemos utilizado, para demostrar los teoremas 2.24.1 y 2.24.2, algunos perfeccionamientos propuestos por A. I. Sajanenko.

§ 26. Se expone una de las variantes del método numérico de Ruffson para determinar el extremo de la función. Véase la exposición con más detalles en [95]. Hemos adoptado el ejemplo 3 del libro de Rao [76].

§ 27. La investigación de la conciliabilidad de la e.v.m. fue comenzada en los años 30 y 40 en los trabajos de Doob [29], Wald [88], Wolfowitz [94] y Cramer [25]. Las principales condiciones de conciliabilidad comprenden, en [88] (además de las condiciones (A_μ), (A_c), (A_0)), la pertenencia de $f_t(x)$ a la clase D_0 y la integrabilidad de

$$\int \ln f_t^{\Delta}(x) f_\theta(x) \mu(dx).$$

En la monografía [48] fueron obtenidas las condiciones de conciliabilidad que utilizan la convergencia

$$\int \sup_{|u| \leq \Delta} (\sqrt{f_{t+u}(x)} - \sqrt{f_t(x)})^2 \mu(dx) \to 0 \quad \text{para} \quad \Delta \to 0.$$

Los resultados de los teoremas 27.1 y 27.2 y de sus corolarios son más generales. El método de demostración es semejante a [88]. La suficiencia de las condiciones (A_0^0) y (2.27.2) fue revelada por A. I. Sajanenko.

§§ 28 y 29. Véanse los comentarios a los §§ 23—27. Hemos adoptado el ejemplo 2.28.1. del libro de Van der Waerden [86].

En la exposición de los párrafos 28 y 29 hemos introducido varios perfeccionamientos en comparación con la variante inicial, o sea, mejoras propuestas por A. I. Sajanenko (en particularidad, hemos añadido el teorema 2.29.5). Estas modificaciones permitieron simplificar el texto en los §§ 13—15 del capítulo 3.

§ 30. La estimación sucesiva se expone con más detalles, por ejemplo, en [95].

§§ 31 y 32. Por lo visto, fue Laplace quien introdujo por primera vez los intervalos confidenciales. Aún en 1812 él mostró que se podía invertir respecto a p la afirmación acerca del grado de divergencia de la frecuencia observada y de la probabilidad binomial p, con el fin de hallar el intervalo para los posibles valores de p. En 1927, Wilson dio la justa interpretación de los intervalos confidenciales (la cual no supone la casualidad del parámetro).

En 1930, Fisher, en [36], propuso un método general de determinación de los intervalos confidenciales exactos. En 1937 y 1938 Neyman desarrolló la teoría general de afirmaciones confidenciales y estableció su relación con la teoría de verificación de las hipótesis. La moderna exposición, muy completa, de esta cuestión se puede hallar en el libro de Lehmann [57]. Hemos utilizado esta exposición en el § 3.7.

El teorema 2.32.1 y el lema 2.23.2 le pertenecen a Fisher.

Capítulo 3

Las primeras aplicaciones de los criterios estadísticos remontan a Laplace (final del siglo 18). El uso sistemático de los criterios para verificar las hipótesis se inicia a partir de los trabajos de K. Pearson, quien propuso, en 1900, el criterio χ^2. Los principales conceptos de errores de primero y segundo género fueron introducidos en 1928 por Neyman y Pearson en [68]. Estos mismos autores fueron los primeros en concebir la importancia de las alternativas para elegir racionalmente el criterio. En la obra conclusiva de Neyman y Pearson [69] se desarrolla la teoría del c. u.m.p.

El libro de Lehmann [57] contiene la exposición sistemática de la teoría de verificación de las hipótesis.

§§ 1—3. El lema fundamental de Neyman — Pearson fue obtenido en [69]. Los teoremas 3.1.1 se pueden extraer del libro de Blackwell Girshik [7]. El libro de Lehmann [57] contiene el teorema 3.2.1. El teorema 3.3.1 de las grandes divergencias le pertenece a Cramer (véase [11]). La estimación de la calidad de los criterios, relacionada con las probabilidades de las grandes divergencias, constituyó la base del concepto de eficacia del criterio de Bahadur. En [3] se exponen los resultados de las investigaciones con arreglo a esta tendencia.

La importancia de la estadística de aportación eficiente fue revelada aún en 1925 en la obra de Fisher [35]. En lo sucesivo, el enfoque relacionado con el estudio de las hipótesis semejantes fue desarrollado intensamente en los trabajos de Le Cam, Roussas y Chíbisov (véanse también los comentarios a los §§ 3.14 y 3.15).

§ 4. La referida concepción general de los criterios estadísticos ha sido universalmente reconocida (véanse [25] y [57]). El concepto de c.u.m.p. fue introducido por Neyman y Pearson en [69]. Aún en el siglo 19, Laplace utilizó el enfoque bayesiano.

§§ 5—8. Los resultados principales de estos párrafos se han tomado del libro de Lehmann [57]. La exposición también es semejante a la de este libro y se distingue por el hecho de que se basa no en el lema generalizado de Neyman — Pearson (lema 3.5.2, véase también [57]), sino en el enfoque bayesiano. Esto simplifica la exposición y la hace más armoniosa.

Ciertas observaciones referentes a los conjuntos confidenciales se exponen en los comentarios a los §§ 2.31 y 3.32.

En el libro de Grenander [39] se examina la posibilidad de extender los resultados principales a los procesos aleatorios.

§ 9. Los autores del teorema 3.9.1 son Hodges y Lehmann [44].

§ 10. El papel fundamental de la relación de verosimilitud en la estadística matemática fue aclarado en los trabajos de Neyman y Pearson [68], [69]. Al estudio del c.r.v. se han dedicado muchos libros. Ciertas tentativas de establecer unas u otras propiedades de optimización asintótica de este criterio se ofrecen en los trabajos [2], [88], [71], [93] y [45].

§ 11. Wald [89] fue quien más contribuyó al desarrollo de la teoría del análisis secuencial. La exposición más completa de los resultados principales, por la cual nos guiamos en nuestro libro, se ofrece en [57].

§ 12. Los criterios de Kolmogórov y ω^2 se exponen en el § 1.8 y en los comentarios a este último. A su vez, algunas modificaciones del criterio de Kolmogórov, que proporcionan la potencia máxima posible, se dan en [16]. El criterio de Moran fue propuesto en [64]. Su potencia para las alternativas semejantes se estudió en [91] y [20].

§ 13. El carácter asintóticamente bayesiano del c.r.v. ue determinado en el trabajo del autor de [10]. Los resultados de la distribución límite de la relación de verosimilitud para la hipótesis principal fueron obtenidos por Wilks [92] y Wald [87] (véase también el libro de Wilks [93]). Wald utilizó la idea de sustituir la hipótesis compleja por una hipótesis promediada. En el trabajo [60] se examina la forma asintótica de los criterios bayesianos. Véanse también los comentarios a los §§ 28 y 29 del capítulo 2.

§§ 14 y 15. Las principales ideas relacionadas con la determinación de los tests asintóticamente óptimos para hipótesis semejantes se exponen en las obras de Wald [87], Le Cam, Roussas (véase el libro de Roussas [81]) y Chíbisov [22]. En el libro [14] se analiza la posibilidad de extender los resultados principales al caso del parámetro de dimensión infinita (es decir, a los procesos aleatorios). La forma de exposición de los §§ 14 y 15 está poco relacionada con los trabajos citados. En el libro [87] de Wald se ofrece la reducción del problema inicial A a un problema B para el parámetro de distribución normal al determinar los criterios óptimos de los principales tipos de problemas examinados en el § 14. La afirmación del teorema 3.15.4 acerca de la distribución de la estadística $2 \ln R_1(X)$ para la hipótesis H_1 se examina en [93]. Véanse también los comentarios a los §§ 28 y 29 del capítulo 2.

§§ 16 y 17. En el año 1900, K. Pearson propuso el criterio χ^2, al cual se han dedicado muchos libros (véase, por ejemplo, la monografía especial de Lancaster [56]). El examen de las diversas propiedades de la optimización se expone en [87], [71], [93], [45], etc. El comportamiento de la potencia del criterio χ^2 al aumentar el número de grupos se analiza, por ejemplo, en [12] y [21]. Los ejemplos 3.16.1 y 3.17.2 se han adoptado del libro de Cramer [25], y el ejemplo 3.17.1, del libro de Rao [76].

§ 18. Al estudiar la estabilidad de las decisiones estadísticas es muy difícil seguir la etapa inicial de ese estudio. Las investigaciones posteriores se basan en los trabajos de Takeuchi, Hodges y Lehmann. En el libro [47] de Huber se hace un resumen detallado de dicha tendencia.

Capítulo 4

§ 1. El criterio χ^2 en el problema del ejemplo 4.1.1, el criterio de Student en el problema del ejemplo 4.1.3 y el criterio de Fisher en los problemas de los ejemplos 4.1.4 y 4.1.5 se utilizan muy a menudo. En el libro [57] de Lehmann se dan otras propiedades de optimización de estos criterios. El ejemplo 4.1.1A se ha tomado del libro [76]. Hay muchos libros (véase [57]) dedicados al problema de Beherns — Fisher (ejemplo 4.1.6).

§ 2. Gnedenko y Koroliuk (véase [32]) hallaron la distribución exacta de la estadística $D_{n,n}$, y Smirnov, la distribución límite de la estadística D_{n_1,n_2}. El teorema 4.2.2. fue demostrado por primera vez en [62] con ayuda del método de momentos. Los criterios de signos y de Wilkoxon también se ofrecen en [41].

§§ 3 y 4. Los problemas de regresión y análisis de varianza se exponen más detalladamente en las monografías especiales de Seber [83] y Scheffe [82]. Véanse asimismo [25], [57] y [76].

§ 5. La observación acerca de la optimización asintótica del criterio (4.5.3) fue tomada de [10].

Capítulo 5

En matemática, la tendencia relacionada con la teoría de los juegos surgió tras la publicación de los trabajos de Borel en 1921 y de von Neumann en 1928. En la estadística matemática, como trabajo inicial, que preparó el uso de la teoría de los juegos, puede considerarse la obra clásica de Neyman y Pearson [70], en la que se enuncian muchas ideas fundamentales de la teoría de las decisiones estadísticas. Wald contribuyó considerablemente al desarrollo de la teoría general de las decisiones estadísticas. En su libro conclusivo [90] se exponen los postulados fundamentales de esta teoría. No obstante, la teoría matemática general de los juegos adquirió su pleno desarrollo en el libro de von Neumann y Morgenstern [65].

Los fundamentos de la teoría de los juegos estadísticos plantean de una forma muy accesible en los libros de Girshik y Blackwell [7] y de Ferguson [33].

§ 2. El libro de McKinsey [63] constituye una introducción relativamente completa a la teoría ordinaria de los juegos.

§§ 3 y 4. En [7] y [33] se da una descripción más completa de los fundamentos de la teoría de los juegos estadísticos. En estos libros, dos teoremas fundamentales de la teoría de los juegos estadísticos sólo se demuestran en el caso particular, para los conjuntos discretos D y θ. Ello se explica por el hecho de que la exposición en el caso general es muy compleja (véase [90]). En el Suplemento VIII se da la demostración más simple que conocemos de tales teoremas, la cual fue hallada por A. I. Sajanenko.

El papel del enfoque bayesiano en distintos tiempos se evaluaba de manera diferente. El mismo ha sido ampliamente utilizado por Laplace en el siglo pasado. Después fue criticado por Fisher, y en los años 20 y 30 de nuestro siglo, el centro de gravedad se desplazó hacia las estimaciones eficientes y asintóticamente eficientes. Más tarde, a medida que se concebía la importancia fundamental del referido enfoque, otra vez comenzó a crecer el interés por él. Esa importancia fundamental es aclarada en los teoremas 5.3.1 y 5.3.2.

§ 5. El concepto fundamental de estadística suficiente fue introducido por R. Fisher [34] en el año 1922. R. Fisher [34] y más tarde J. Neyman [66] propusieron un criterio simple que revela la existencia y el tipo de estadística suficiente. Este criterio es conocido con el nombre de teorema de factorización de Neyman — Fisher y está representado en el teorema 2.12.1. La estricta demostración del teorema de Neyman — Fisher, desde el punto de vista de la teoría de los conjuntos, fue obtenida tan sólo en 1949 por Halmos y Savage [43].

El concepto de σ-álgebra suficiente es más amplio que el concepto de estadística suficiente. Las condiciones necesarias y suficientes para su coincidencia se dan en [95]. El teorema 5.5.1 (primero para la función cuadrática de pérdidas) fue independientemente obtenido por Blackwell *[6] (1947)*, Rao *[74] (1945)*, *[75] (1949)* y *Kolmogórov [53] (1950)*. Las generalizaciones para el caso de función arbitraria de pérdidas están íntimamente ligadas a los nombres de Lehmann y Scheffe [95].

A Hotelling y Pitman les pertenece la idea de utilizar las consideraciones invariantes. Ch. Stein (véanse [95] y [48]) contribuyó considerablemente al desarrollo de la teoría.

En [95] se ofrecen datos más detallados acerca del carácter no desplazado.

§ 6. El libro [48] de Ibraguímov y Jasminski contiene resultados semejantes a los teoremas de este párrafo.

§ 7. El carácter asintóticamente bayesiano del c.r.v. fue establecido en el trabajo del autor de [10]. Los resultados de la distribución límite de la relación de verosimilitud para la hipótesis principal fueron obtenidos por Wilks [92] y Wald [87] (véase también el libro de Wilks [93]). Wald utilizó la idea de sustituir la hipótesis compleja por una hipótesis promediada. El tipo asintótico de criterios bayesianos se expone en [60].

§ 8. Las principales ideas relacionadas con la determinación de los tests asintóticamente óptimos para hipótesis semejantes se examinan en los trabajos de Wald [87], Le Cam, Roussas (véase el libro [81] de Roussas) y Chíbisov [22]. En [15] se estudia la posibilidad de extender los resultados principales al caso de un parámetro de dimensión infinita (es decir, a los procesos aleatorios). La forma de exposición del § 8 y de los §§ 14 y 15 del capítulo 3 está poco relacionada con los trabajos citados. La reducción del problema inicial A a un problema B (para el parámetro de distribución normal), al determinar los criterios óptimos para los principales tipos de problemas, se analiza en el trabajo de Wald [87].

Suplemento VII

Fue A. A. Mogulski quien propuso utilizar el lema 1 para demostrar el teorema 2.28. La demostración de este lema se remonta a S. L. Sóbolev. La demostración del lema 1 también se puede obtener fácilmente utilizando los resultados de [96]. En la edición rusa del libro se da otra demostración del teorema 2.28, la cual utiliza ciertas ideas de A. N. Kolmogórov acerca de la estimación de la distribución del máximo del proceso aleatorio.

Suplemento VIII

La demostración de dos teoremas fundamentales de la teoría de los juegos estadísticos se ofrece en [90] y, para suposiciones más particulares, en [7] y [33]. En el libro presente se expone el enfoque de la demostración propuesta por A. I. Sajanenko. Su parte central consta de los lemas 2 y 3. De hecho, el lema 2 no está relacionado con el carácter estadístico del juego, se basa en los teoremas de Hahn — Banach y de Riss y por su idea se asemeja a los razonamientos utilizados, por ejemplo, en [31]. La demostración del lema 3 se basa en los teoremas de Kolmogórov [54] y Prójorov [5].

Al trazar las tablas I—IV se utilizó el libro de Bolshev y Smirnov [8].

Bibliografía

1. *Bahadur R. R.* On Fisher's bound for asymptotic variances. Ann. Math. Statist., 1964, 35, 4, 1545—1552.
2. *Bahadur R. R.* An optimal property of the likelihood ratio statistic, Proc. 5-th Berkeley Sympos. Math. Statist. Prob. — Berkeley — Los Angeles, v. 1, 1965, 27—40.
3. *Bahadur R. R.* Some limit theorems in statistics. — Philadelphia: S.I.A.M., 1971.
4. *Bahadur R. R., Lehmann E. L.* Two comments on "Sufficiency and statistical decision functions". — AMS. 1955, 26, 139—141.
5. *Billingsley P.* Convergence of probability measures, N.Y., Wiley, 1968.
6. *Blackwell D.* Conditional expectation and unbiased sequential estimation. — Ann. Math. Statist., 1947, 18, 105—110.
7. *Blackwell D., Girshik M. A.* Theory of games and statistical decisions, N.Y., Wiley, 1954.
8. *Большев Л. Н., Смирнов Н. В.* Таблицы математической статистики. — М.: Наука, 1965.
   (*Bólshev L. N., Smirnov N. V.* Tablas de estadística matemática.)
9. *Боровков А. А.* Вероятностные процессы в теории массового обслуживания. — М.: Наука, 1972.
   (*Borovkov A. A.* Procesos probabilísticos en la teoría de las colas.)
10. *Боровков А. А.* Асимптотически оптимальные тесты для проверки сложных гипотез. — Теория вероятн. и ее примен., 1975, 20, 3, 463—487.
   (*Borovkov A. A.* Tests asintóticamente óptimos para verificar las hipótesis compuestas. — Teoría de las probabilidades y su aplicación.)
11. *Боровков А. А.* Теория вероятностей. — М.: Наука, 1976,
   (*Borovkov A. A.* Teoría de las probabilidades.)
12. *Боровков А. А.* О мощности критерия при увеличении числа групп. — Теория вероятн. и ее примен., 1977, 22,2, 375—379.
   (Acerca de la potencia del criterio al aumentar el número de grupos. Teoría de las probabilidades y su aplicación.)
13. *Боровков А. А., Саханенко А. И.* Неравенства типа Рао-Крамера для байесовского риска. — Теория вероятн. и ее примен. 1980, 25, 1, 207—209.
   (*Borovkov A. A., Sajanenko A. I.* Desigualdades del tipo de Rao — Cramer para el riesgo bayesiano.)
14. *Боровков А. А., Саханенко А. И.* Об асимптотически оптимальных тестах для проверки сложных гипотез. — Труды Института математики СО АН СССР, 1981, т. 1.
   (Acerca de los tests asintóticamente óptimos para verificar las hipótesis compuestas.)

15. *Боровков А. А., Саханенко А. И.* Об асимптотически оптимальных тестах для проверки сложных близких гипотез. — Труды Института математики СО АН СССР, 1982, т. 1.
(*Borovkov A. A., Sajanenko A. I.* Acerca de los tests asintóticamente óptimos para verificar las hipótesis compuestas semejantes.)
16. *Боровков А. А., Сычева Н. М.* О некоторых асимптотически оптимальных непараметрических критериях. — Теория вероятн. и ее примен., 1968, 13, 3, 385—418.
(*Borovkov A. A., Sicheva N. M.* Acerca de algunos criterios no paramétricos asintóticamente óptimos.)
17. *Chapman D. G., Robbins H. E.* Minimum variance estimation without regularity assumptions. — Ann. Math. Statist., 1951, 22, 581—586.
18. *Ченцов Н. Н.* Об оценке неизвестного многомерного нормального распределения. — Теория вероятн. и ее применен., 1967, 12, 4, 619—633.
(*Chentsov N. N.* Acerca de la estimación de la distribución normal multidimensional media desconocida. Teoría de las probabilidades y su aplicación.)
19. *Ченцов Н. Н.* Статистические решающие правила и оптимальные выводы. — М.; Наука, 1972.
(*Chentsov N. N.* Reglas estadísticas de decisión y deducciones óptimas.)
20. *Чибисов Д. М.* О критериях согласия, основанных на выборочных промежутках. — Теория вероятн. и ее примен., 1961, 6, 1, 354—358.
(Chíbisov D. M. Acerca de los criterios de aceptación basados en los intervalos muestrales. — Teoría de las probabilidades y su aplicación.)
21. *Чибисов Д. М., Гванцеладзе Л.* О критериях согласия, основанных на сгруппированных данных. — В кн,: III Советско-японский симпозиум по теории вероятностей. — Ташкент: Фан, 1975, 183—185.
(*Chíbisov D. M., Gvantseladze L.*. Acerca de los criterios de aceptación basados en los datos agrupados.)
22. *Chíbisov D. M.* Transition to the limiting process for deriving asymptotically optimal tests. Sankhya, 1969, A31, 3, 241—258.
23. *Cox D., Hinkly D.* Theoretical statistics, Chapman and Hall, London, 1974.
24. *Cramer H.* A contribution to the theory of statistical estimation. — Aktuariestidskrift, 1946, 29, 458—463.
25. *Cramer H.* Mathematical Methods of Statistics, 1946.
26. *David H. A.* Order Statistics, N.Y., Wiley, 1976.
27. *De Hardt T.* Generalizations of the Glivenko — Cantelli theorem. — Ann. Math. Stat., 1971, 42, 2050—2055.
28. *Donsker M.* Justifications and extension of Doob's heuristic approach to the Kolmogórov — Smirnov theorems. — Ann. Math. Statist., 1952, 23, 277—281.
29. *Doob J. L.* Probability and statistics. — Trans. Amer. Math. Soc., 1934, 36, 4, 759—775.
30. *Doob J. L.* Stochastic Processes, N.Y., Wiley, 1953.
31. *Edwards R. E.* Functional analysis, HOLT, Rinehart and Winston, New York, Chicago, San Francisco, Toronto, London, 1965.
32. *Feller W.* An Introduction to Probability Theory and its Applications, vols. I, II, N.D., Wiley Eastern, 1972.
33. *Ferguson J. S.* Mathematical statistics. A decision theoretic approach. — New York and London: Academic Press, 1967.
34. *Fisher R. A.* On the mathematical foundations of theoretical statistics. — Phil. Trans. Roy. Soc. A, 1922, 222, 309—368.
35. *Fisher R. A.* Theory of statistical estimation. — Proc. Camp. Phil. Soc., 1925, 22, 700—725.

36. *Fisher R. A.* Inverse probability. — Proc. Cambridge Phill. Soc., 1930, 26, 528—535.
37. *Frechet M.* Rev. Intern. de Stat. 1943, 182.
38. *Гихман И. И., Скороход А. В.* Введение в теорию случайных процессов. — М.: Наука, 1977.
(*Guijmán I. I., Skorojod A. V.* Introducción a la teoría de los procesos aleatorios, en ruso.)
39. *Grenander U.* Stochastic processes and statistical interference, Ark. Math., 1,3, 1960, 195—277.
40. *Гусев С. И.* Асимптотические разложения, связанные с некоторыми статистическими оценками в гладком случае. — Теория вероятн. и ее примен. — 1976, 21, 1, 16—33.
(*Gúsev S. I.* Desarrollos asintóticos relacionados con algunas estimaciones estadísticas en un caso suave. — Teoría de las probabilidades y su aplicación.)
41. *Hajek J., Sidak Z.* Theory of rank tests. Academía Publishing House of the Czechoskovak Academy of Sciences, Prague, 1967.
42. *Halmos P. R.* Measure Theory, Princeton, N.Y., Van Nostrand, 1962.
43. *Halmos P. R., Savage L. J.* Application of the Radon-Nikodym theorem to the theory of sufficient statistics. — Ann. Math. Statist., 1949, 20, 225—241.
44. *Hodges J., Lehmann E.* Some problems in minimax estimation. — Ann. Math. Statist., 1950, 21, 2, 182—197.
45. *Hoeffding W.* Asymptotically optimal test for multinomial distributions. — Ann. Math. Statist., 1965, 36, 2, 369—401.
46. *Hotelling H.* The generalization of student's ratio. Ann. Math. Statist., 1931, 2, 360—378.
47. *Huber P. J.* Robust statistics: a review. — Ann. Math. Statist., 1972, 43, 1041—1067.
48. *Ибрагимов И. А., Хасьминский Р.* Асимптотическая теория оценивания. — М.: Наука, 1979.
(*Ibraguímov I. A., Jasminski P.* Teoría asintótica de la estimación.)
49. *Kendall H. G.* Stuart A. Interference and Relationship, Charles Griffin & Company Limited, London, 1967.
50. *Kendall M. G., Stuart A.* The advanced theory of statistics, vol. 2, Ch. Griffin & Company Limited, London, 1961.
51. *Kiefer J.* On minimum variance estimators. — Ann. Math. Statist., 1952, 23, 627—629.
52. *Kiefer J.* On large deviations of the identically distributed functions of vector chance variables and LIL, — Pacif. J. Math. 1961, 11, 2, 649—660.
53. *Колмогоров А. Н.* Несмещенные оценки. — Изв. АН СССР, сер. мат., 1950, 303.
(*Kolmogórov A. N.* Estimaciones no desplazadas.)
54. *Колмогоров А. Н.* Основные понятия теории вероятностей — М.: Наука, 1974.
(*Kolmogórov A. N.* Conceptos fundamentales de la teoría de las probabilidades.)
55. *Kullback S., Leibler R. A.* On information and sufficiency. — Ann. Math. Statist., 1951, 22, 79—86.
56. *Lancaster H. O.* The chi-squared distribution. — N.Y., Wiley, 1969.
57. *Lehmann E. L.* Testing statistical hypotheses, John Wiley, New York, 1959.
58. *Lehmann E. L.* Theory of point estimation, Wiley, N.Y., 1983.
59. *Lehmann E. L., Scheffe H.* Completeness, similar regions and unbiased estimatiion, — Pt. I. Sankhya, 1950, 10, 305—340.
60. *Lindley D.* The use of prior probability distributions in statistical interference and decision. — Proc. 4-th Berkeley Sympos. Math. Statist. Prob., Berkeley — Los Angeles, v. 1, 1960, 453—468.
61. *Loeve M.* Probability Theory, 2nd ed., D. Van Nostrand Co., 1960.
62. *Mann H. B., Whitney D. R.* On a test whether one of two random variables is stochastically lerger than the other. — Ann. Math. Statist. 1947, 18, 50.

63. *McKinsey J. C. C.* Introduction to the theory of games. — McGraw-Hill, N.Y., 1952.
64. *Moran P. A. P.* The random division of an interval, — J. Roy. Stat. Soc., Suppl., 1947, 9, 92—98.
65. *Von Neumann.* Morgenstern, Theory of Games and Economic Behaviour, Princeton, Princeton University Press, 1953.
66. *Neyman J.* Su un teorema concernente le cosidette statistiche sufficienti. — Inst. Ital. Atti. Giorn., 1935, 6, 320—334.
67. *Neyman J.* First course in probability and statistics,, Holt, Rinehart and Winston, JNC, N.Y., 1950.
68. *Neyman J., Pearson E. S.* On the use and interpretation of certain test criteria. — Biometrica, 1928, 20A, 175—240, 263—294.
69. *Neyman J., Pearson E. S.* On the problem of the most efficient test of statistical hypotheses, — Phil. Trans. Roy. Soc., Ser. A, 1933, 231, 289—337.
70. *Neyman J., Pearson E. S.* The testing of statistical hypotheses in relation to probabilities a priori. — Proc. Camb. Phil. Soc. 1933, 24, 492—510.
71. *Oosterhoff J., W. R. van Zwet.* The likelihood ratio test for the multinomial distribution. Proc. 6-th Berkeley Sympos. Math. Statist. Prob., Berkeley — Los Angeles, v. 1, 1970, 31—50.
72. *Parzen E.* On estimation of a probability density function and mode. — Ann. Math. Statist., 1962, 33,3, 1065—1076.
73. *Pitman E. J. G.* The estimation of the location and scale parameters of a continuous population of any given form. Biometrica, 1938, 30, 391—421.
74. *Rao C. R.* Information and accuracy attainable in estimation of statistical parameters. Bull. Calcutta Math. Soc., 1945, 37, 81—91.
75. *Rao C. R.* Sufficient statistics and minimum variance estimates. — Proc. Cambr. Phill. Soc., 1949, 45, 213—218.
76. *Rao C. R.* Linear Statistical Interference and its Applications. 2nd ed., Wiley: N.Y., 1973.
77. *Скороход А. В.* Случайные процессы с независимыми приращениями. — М.: Наука, 1964.
(*Skorojod A. V.* Procesos aleatorios con incrementos independientes.)
78. *Смирнов Н. В.* О распределении ω^2-критерия Мизеса. — В кн.: Смирнов Н. В. Теория вероятностей и математическая статистика. Избранные труды. — М.: Наука, 1970.
(*Smirnov N. V.* Acerca de la distribución del ω^2-criterio de Mises. — En el libro: Smirnov N. V. Teoría de las probabilidades y estadística matemática.)
79. *Rosenblatt M.* Remarks on some nonparametric estimates of a density function. — Ann. Math. Statist., 1956, 27, 3, 832—837.
80. *Rosenblatt M.* Curve estimation. — Ann. Math. Statist., 1971, 42, 6, 1815.1842.
81. *Roussas G.* Contiguity of probability measures. Cambridge University Press, 1972.
82. *Scheffe H.* The analysis of variance, Wiley; N.Y., 1959.
83. *Seber G. A.* Linear regression analysis, Wiley; N.Y., 1977.
84. *Ширяев А. Н.* Вероятность. — М.: Наука, 1980.
(*Shiriaev A. N.* Probabilidad, en ruso.)
85. *Sídorov V. A.* y otros. Measurement of the $\varphi \to \pi^+\pi^-$ branching ratio, — Physics Letters, 1981, 99B, 1, 62—65.
86. *van der Waerden.* Mathematische Statistik. Springer-Verlag, 1957.
87. *Wald A.* Tests of statistical hypotheses concerning several parameters when the number of observations is large. — Trans. Amer. Math. Soc., 1943, 54, 3, 426—482.
88. *Wald A.* Note on the consistency of the maximum likelihood estimate. — Ann. Math. Statist., 1949, 20, 595—601.
89. *Wald A.* Sequential analysis, Wiley; N.Y., 1947.

90. *Wald A.* Statistical decision functions, New York, 1950.
91. *Weiss L.* The asymtotic power of certain tests of fit based on sample spacings. — Ann. Math. Statist., 1957, 28, 3, 783—786.
92. *Wilks S. S.* The large sample distribution of the likeli-hood ratio for testing composite hypothesis. — A. Math. Statist., 1938, 9, 60—62.
93. *Wilks S. S.* Mathematical Statistics, Wiley, N.Y., 1962.
94. *Wolfowitz J.* On Wald's proof of the consistency of the maximum likelihood estimate. — Ann. Math. Statist., 1949, 20, 601—602.
95. *Zacks S.* The theory of statistical interference, Wiley, N.Y., 1971.
96. *Фаддеев Д. К., Вулих Б. З.* и др., Избранные главы анализа и высшей алгебры., Ленинград, Изд-во ЛГУ, 1981, 199 с.
(*Faddéev D. K., Vúlij B. Z.* y otros. Capítulos escogidos del análisis y el álgebra superior.)

Designaciones principales

Las designaciones se dan en orden alfabético: primero el alfabeto ruso, después el latino y el griego. Al final se ofrecen los símbolos matemáticos.

(A_0), condición de correspondencia biunívoca entre el conjunto paramétrico Θ y la familia de distribuciones $\mathscr{P} = \{\mathbf{P}_\theta\}_{\theta\in\Theta}(\mathbf{P}_{\theta_1} \neq \mathbf{P}_{\theta_2}$ si $\theta_1 \neq \theta_2)$

(A_c), condición consistente en que el conjunto paramétrico Θ es compacto

(A_μ), condición en virtud de la cual todas las distribuciones de la familia $\mathscr{P} = \{\mathbf{P}_\theta\}$ son dominadas por la medida μ (existe la densidad $f_\theta = d\mathbf{P}_\theta/d\mu$)

b, $b(\theta)$, desplazamiento

$\mathfrak{B}$, σ-álgebra de los conjuntos de Borel sobre la recta R

$\mathfrak{B}_{\mathscr{X}}$, σ-álgebra en el espacio de fase $\mathscr{X}$ (de los conjuntos de Borel si $\mathscr{X} = R^m$

$\mathbf{B}_p$, distribución polinomial (incluyendo la distribución de Bernoulli)

$C(a, b)$, espacio de las funciones continuas en $[a, b]$.

c.a.b., criterio asintóticamente bayesiano

c.a.u.m.p., criterio asintóticamente uniforme más potente

c.d., casi por doquier

c.m.p., criterio más potente

c.r.v., criterio de la relación de verosimilitud

c.t., casi todos (los)

c.u.m.p., criterio uniformemente más potente

$D(a, b)$, espacio de las funciones en $[a, b]$, continuas a la izquierda (en el punto a a la derecha) y que sólo tienen un número finito de saltos

D, espacio de las estrategias del primer jugador (en el cap. 4)

D_θ, varianza de la distribución $\mathbf{P}_\theta$

$\mathscr{D}$, espacio de las funciones de decisión en un juego estadístico

E, matriz unidad

e.m.c., esperanza matemática condicional

e.v.m., estimación de la verosimilitud máxima

$\mathscr{E}$, familia exponencial de las distribuciones

$f_\theta(x)$, densidad de la distribución $\mathbf{P}_\theta$ respecto a la medida μ

$f_\theta(X)$, función de verosimilitud igual (por definición) a $\prod\limits_{i=1}^{n} f_\theta(x_i)$

$F(x)$, por regla general, la función de distribución correspondiente a la distribución $\mathbf{P}$

$F_n^*(x)$, función empírica de distribución

$F_{k_1k_2}$, distribución de Fisher

G, grupo de transformaciones de $\mathscr{X}$ en sí, correspondiente a la familia invariante
h_ε, cuantila de la distribución χ^2
H_i, hipótesis
H_k, distribución χ^2
I_x, distribución concentrada en el punto x
$I(\theta) = \|I_{ij}(\theta)\|$, $I_{ij}(\theta) = \mathbf{M}_\theta \frac{\partial}{\partial\theta_i} l(x_1, \theta) \frac{\partial}{\partial\theta_j} l(x_1, \theta)$, matriz de información de Fisher
I_A, indicador del conjunto A
K_b, clase de estimaciones con desplazamiento $b = b(\theta)$
K_0, clase de estimaciones no desplazadas
$\tilde{K}_0$, clase de estimaciones asintóticamente no desplazadas
$\tilde{K}^0$, clase de estimaciones asintóticamente centrales
$K_{\Phi,2}$, clase de estimaciones asintóticamente normales θ^*, para las cuales $\mathbf{M}_\theta n(\theta^* - \theta)^2 \to \sigma^2(\theta)$, donde $\sigma^2(\theta)$ es la varianza de la distribución normal límite para $\sqrt{n}(\theta^* - \theta)$
K_ε, (en el cap. 3) clase de criterios de dimensión ε (de nivel $1 - \varepsilon$)
K_ε, clase de criterios no desplazados de dimensión ε
$\tilde{K}_\varepsilon$, clase de criterios de nivel asintótico $1 - \varepsilon$
$K_\varepsilon^{Q_1}$, clase de criterios de dimensión ε para el enfoque parcialmente bayesiano
$\tilde{K}_\varepsilon^{Q_1}$, clase de criterios de dimensión asintótica ε para el enfoque parcialmente bayesiano
$K_{\alpha_1,\dots,\alpha_{r-1}}$, clase de criterios con valores fijos α_i de las probabilidades de los errores de i-ésimo género, $i = 1, \dots, r - 1$
$K_{\alpha,\sigma}$, distribución de Cauchy
$l(x, \theta) = \ln f(x)$
$L(X, \theta) = \ln f_\theta(X)$, función logarítmica de verosimilitud
L_{α,σ^2}, distribución lognormal
$\mathbf{M}_\theta$, esperanza matemática de la distribución $\mathbf{P}_\theta$
$\mathbf{M}(\xi/\mathfrak{U})$, esperanza matemática condicional ξ respecto a la σ-álgebra $\mathfrak{U}$
$\mathbf{M}(\xi/\eta)$, esperanza matemática condicional ξ respecto a la variable aleatoria η
n, volumen de la muestra
N_P, N_F, portador de la distribución $\mathbf{P}$ con la función de distribución F
$\mathbf{P}$, símbolo de la distribución, utilizado en distintos sentidos
$P(B/y)$, distribución condicional
$\mathbf{P}_n^*$, distribución empírica
$\mathbf{P}_\theta$, distribución dependiente del parámetro
$\mathscr{P}$, familia de distribuciones
$\mathbf{Q}$, estrategia randomizada de la "naturaleza" (distribución a priori de θ)
$\mathbf{Q}_x$, distribución a posteriori de θ
$\overline{\mathbf{Q}}$, la peor distribución de θ (estrategia minimáx de la "naturaleza")
$q(t/X)$, densidad de distribución a posteriori de θ
R, recta real
R^m, espacio euclídeo m-dimensional
(R), condición de regularidad de la familia paramétrica en cuya virtud la función $\sqrt{f_\theta(x)}$ es continuamente derivable respecto a θ, y la información de Fisher es positiva y continua
(RR), condiciones de regularidad de la familia paramétrica, que exigen el cumplimiento de las condiciones (A_0), (A_c) y (R), así como de la derivabilidad continua de segundo orden de la función $l(x, \theta)$ y de la existencia de la mayorante $l(x) \geqslant |l''(x, t)|$, para la cual la integral de $\mathbf{M}_\theta l(x_1)$ converge uniformemente hacia Θ
$S = S(X)$, estadística
S^2, varianza empírica
S_X^2, varianza empírica correspondiente a la muestra X

$S_0^2 = \dfrac{n}{n-1} S^2$

$\mathbf{T}_k$, distribución de Student

$\mathbf{U}_{a,b}$, distribución uniforme en $[a, b]$

$u^* = \sqrt{n}(\bar{\theta}^* - \theta)$, estimación normalizada de verosimilitud máxima

$w(t)$, (no siempre) proceso wieneriano

$w^0(t)$, puente browniano

$w^n(t)$, proceso empírico

x_i — elemento de la muestra

$X = X_n = (x_1, \ldots, x_n)$ — muestra de volumen n

$[X_\infty]_n = X_n$ — parte de una muestra infinita, constituida por primeros h elementos de esta última

$x_{(i)}$, i-ésimo elemento de una serie variacional

$\bar{x}$, media empírica

$\mathscr{X}$, espacio al cual pertenecen observaciones (espacio de fase de la muestra)

$(\mathscr{X}, \mathfrak{B}_{\mathscr{X}}, P)$, espacio probabilístico muestral correspondiente a una observación

$(\mathscr{X}^n, \mathfrak{B}^n_{\mathscr{X}}, P)$, espacio probabilístico muestral correspondiente a la muestra de volumen n

$x = (x_1, \ldots, x_n)$, elemento de $\mathscr{X}^n$

$\alpha_i(\pi)$ — probabilidad del error de i-ésimo género del criterio π

$\beta(\delta)$, potencia del criterio δ

$\beta_\pi(\delta)$, función de potencia del criterio π

$\beta_{\lambda_1,\lambda_2}$, distribución beta

$\Gamma_{\alpha,\lambda}$, distribución gamma

$\delta = \delta(X)$, (en el cap. 3) regla (criterio) de decisión o (en cap. 5) función de decisión

δ, estrategia del primer jugador

ζ_p, cuantila de orden p

ζ_p^*, cuantila muestral de orden p

θ, parámetro (estrategia de la "naturaleza")

$\theta^{\pm}$, fronteras del intervalo confidencial para el parámetro θ

θ^*, estimación del parámetro θ

θ_Q^*, estimación bayesiana del parámetro θ, la cual corresponde a la distribución a priori Q

$\bar{\theta}^*$, estimación minimáx del parámetro θ

$\hat{\theta}^*$, estimación de verosimilitud máxima del parámetro θ

Θ, conjunto de valores posibles del parámetro θ

Θ^*, conjunto confidencial

λ_ε, cuantila de la distribución normal

$\pi = \pi(X)$, (en el cap. 3) criterio randomizado o (en los caps. 3 y 5) regla (criterio) randomizada de decisión

π, estrategia randomizada del primer jugador

π_Q, criterio (estrategia) bayesiano correspondiente a la distribución a priori Q

$\pi_{Q_1Q_2}$, criterio bayesiano para el enfoque parcialmente bayesiano

$\bar{\pi}$, criterio (estrategia) minimáx

$\hat{\pi}$, criterio de la relación de verosimilitud

π^0, criterio uniformemente más potente

Π_λ, distribución de Poisson

Φ_{α,σ^2}, distribución normal

$\Phi(x)$, función de la distribución estándar normal

$\underset{d}{=}$, símbolo que significa la coincidencia de las distribuciones de muestras o de variables aleatorias

$\xrightarrow[P]{}$, signo de convergencia en probabilidad

$\xrightarrow[c.s.]{}$, signo de convergencia casi segura (con probabilidad 1)

$\Rightarrow$, signo de convergencia débil de las distribuciones (se utiliza tanto entre las variables aleatorias como entre las distribuciones)

$\in$, signo utilizado entre las designaciones de la muestra (de la variable aleatoria) y de la distribución: significa que la muestra fue extraída de una distribución dada (la variable aleatoria tiene una distribución dada)

$\Subset$, signo de convergencia débil. La relación $\xi_n \Subset \mathbf{P}$ quiere decir que la distribución ξ_n converge débilmente hacia $\mathbf{P}$ cuando $n \to \infty$

Indice alfabético de materias

A NUESTROS LECTORES:

Mir edita libros soviéticos traducidos al español, inglés, francés, árabe y otros idiomas extranjeros. Entre ellos figuran las mejores obras de las distintas ramas de la ciencia y la técnica; manuales para los centros de enseñanza superior y escuelas tecnológicas, literatura sobre ciencias naturales y médicas. También se incluyen monografías, libros de divulgación científica y ciencia-ficción.

Dirijan sus opiniones a la Editorial Mir, 1 Rizhski per., 2, 129820, Moscú, I-110, GSP, URSS.